THE UNIVERSITY OF ALBERTA

RELATION OF SELECTED APTITUDES TO THE VERBAL AND DIAGRAMMATIC CONTENT OF PHYSICS TESTS

by

© John D. George

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF MASTER OF EDUCATION

Department of Secondary Education

Edmonton, Alberta

Fall, 1971

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled "Relation of Selected Aptitudes to the Verbal and Diagrammatic Content of Physics Tests" submitted by John D. George, in partial fulfilment of the requirements for the degree of Master of Education.

ABSTRACT

The central idea for the study evolved from the observation that the best single predictor of physics achievement was verbal reasoning. This observation appeared to be anomalous in view of the visual nature of many fundamental ideas in physics. It appeared reasonable that one method of limiting the verbal aspect of test items could be the use of diagrams in the physics tests and that such tests might correlate more closely with the spatial or mechanical reasoning.

Five tests of achievement in grade eleven PSSC physics were constructed of which the first was completely verbal (PhV); the second test was numerical, requiring a calculation(s); the third was based on the interpretation of graphs usually requiring some numerical calculation; the fourth test included a diagram, although the items contained a verbal stem and distractors, and the fifth test contained items in which both the stem and the distractors were diagrammatic in form.

The Verbal Reasoning (VR), Numerical Ability (NA), Abstract Reasoning (AR), Mechanical Reasoning (MR), and Space Relations (SR) tests of the Differential Aptitude Test battery were administered to a sample of seventy-four Grade XI PSSC Physics students at Vincent Massey Collegiate in Winnipeg. The students also wrote the five physics tests.

Analysis of the results of the DAT revealed that the students in the sample were generally superior to the standardizing population for the DAT norms, particularly in the VR, NA, and AR subtests. The DAT subtests were not independent since the amount of shared variance varied from eight per cent to forty-seven per cent.

Reliability coefficients, levels of difficulty, and biserial correlation coefficients were determined for each of the five physics tests. The K-R 20 Formula reliability coefficients were generally low (from .45 to .68). This was regarded as partly due to the relatively small number of items in each of the tests (from ten to seventeen). Extension of the tests to approximately thirty items using the Spearman-Brown prophecy formula yielded reliability coefficients ranging from .71 to .81 for the extended tests, to justifying their use for further analysis.

Intercorrelations among the five physics tests were determined and revealed a rather large amount of shared variance (from eight per cent to forty-three per cent) among the physics tests.

Correlation coefficients for the DAT subtests and the five physics tests were determined and significant ($p < 0.05$) correlations were observed between all but two of the DAT vs. physics tests.

Stepwise Regression Analysis revealed that, for the verbal physics test (PhV), the significant ($p < 0.05$) predictors were VR and NA; for the numerical physics test (PhN), the significant predictors were VR, NA, and SR; for the graphic physics test (PhG), the significant predictors were VR, NA, AR, and MR; for the diagrammatic-verbal physics test (PhDV), the significant predictors were VR and SR, and for the pure diagrammatic physics test the only significant predictor was MR.

The appearance of VR as the first significant predictor in the first four multiple regression equations was partly accounted for by

the measurement of a general reasoning ability by VR. There was a notable and measureable difference between the performance of students on the pure diagrammatic test and on the other physics tests.

It may be important that teachers, in designing tests of achievement in physics, keep in mind the various aptitudes which students have. Students with high mechanical aptitude may be able to identify the correct solution to a problem when it is expressed in a pure diagrammatic form without the complex verbal aspect often associated with physics problems. It appears that only if the stem and distractors are diagrammatic with a minimum of verbal content is the verbal reasoning factor not significant.

ACKNOWLEDGEMENTS

I would like to acknowledge the very considerable debt I owe to Dr. Heidi Kass and Dr. Marshall Nay who were my advisers at various times during the course of this investigation. Their help and advice were particularly necessary in view of the fact that I was trying to complete the thesis in Winnipeg.

The Research and Planning Branch of the Department of Youth and Education for Manitoba and Mr. Mike Yakimishyn in particular were most helpful to me in obtaining the 'Stepwise Regression' program from the University of Alberta; in assisting me in its use, and in providing computer time at the University of Manitoba.

The Fort Garry School Division also assisted materially in providing facilities for scoring the tests, for which I am grateful.

I would like to thank Dr. W. Brouwer and Professor N. M. Purvis for their encouragement and for consenting to act on the committee.

Finally, my thanks to my wife and family who have been very patient.

TABLE OF CONTENTS

CHAPTER		PAGE
I	INTRODUCTION	1
	I. The Problem	1
	II. Justification of the Study	3
	III. Definitions and Delimitations	5
	IV. Organization of the Thesis	7
II	REVIEW OF THE LITERATURE	9
	I. The Structure of the Human Intellect	9
	II. Related Studies	15
	III. Summary	22
III	THE TESTS, STATISTICAL PROCEDURES AND HYPOTHESES	24
	I. The Tests	24
	II. Statistical Procedures	41
	III. Hypotheses	45
IV.	RESULTS AND DISCUSSION	48
	I. Analysis of the Results for the Differential Aptitude Tests	48
	II. Analysis of the Results of the Physics Tests	54
	III. Intercorrelations Among the Physics Tests and the DAT	70
	IV. The Stepwise Regression Analysis	75
V	SUMMARY AND RECOMMENDATIONS	91
	I. Summary	91

TABLE OF CONTENTS (cont'd.)

CHAPTER			PAGE
V	II.	Recommendations.	95
	III.	Recommendations for Further Research	96
BIBLIOGRAPHY.			98
APPENDIX.			

LIST OF TABLES

TABLE		PAGE
I	Percentile Scores for the DAT -- Differentiated by Sex.	49
II	The Means, Standard Deviations, and Intercorrelations for the DAT Subtests	52
III	Level of Difficulty, Biserial Correlation and Reliability Coefficients for Five Physics Tests.	55
IV	The Means, Standard Deviations, and Intercorrelations for the Physics Tests.	68
V	Intercorrelations Among the DAT Subtests and the Physics Tests.	71
VI	Stepwise Prediction of Physics-Verbal Scores (PhV) from the DAT Subtests.	77
VII	Stepwise Prediction of Physics-Numerical Scores (PhN) from the DAT Subtests.	79
VIII	Stepwise Prediction of Physics-Graph Scores (PhG) from the DAT Subtests.	81
IX	Stepwise Prediction of Physics-Diagrammatic Verbal Scores (PhDV) from the DAT Subtests.	84
X	Stepwise Prediction of Physics-Pure Diagrammatic Scores (PhPD) from the DAT Subtests.	86
XI	Significant DAT Predictors of Physics Test Scores	88
XII	Item Analysis of the Physics Tests.	

LIST OF FIGURES

FIGURE		PAGE
1	The structure of educational abilities.	11
2	General plan of the relationship among the various main categories of intellectual abilities.	12
3	Frequency distribution of scores for PhV.	59
4	Frequency distribution of scores for PhN.	60
5	Frequency distribution of scores for PhG.	61
6	Frequency distribution of scores for PhDV	62
7	Frequency distribution of scores for PhPD	63

CHAPTER I

INTRODUCTION

I. THE PROBLEM

One aspect of the problem of evaluation of achievement in physics is that of the bias inherent in the tests and examinations which are used. It has been shown in studies which will be reviewed later in Chapter II, Review of the Literature, that the best predictor of success in physics tests has been a verbal reasoning and numerical ability test. It is the position of the investigator that teachers of physics may be discriminating against some students because of the type of physics test items which are used in high school tests. The usual test or examination in physics is composed of items in which the verbal content is substantial and usually somewhat complex. If it were possible to design items which would test the same concepts efficiently, but with a lesser emphasis on verbal skills, it might prove to be advantageous to students whose verbal skills are not highly developed.

Studies which have been conducted on the efficiency of prediction of intelligence tests and aptitude tests for success in school science, and particularly in physics, have used the usual kind of school examination as the criterion of success in physics. Of interest in this study are the results which would be achieved if other kinds of instruments of evaluation were used. If one were to vary the verbal content of the test items, and to emphasize the visual nature of physics in test items, would the mechanical-spatial factor become a

more efficient predictor of success?

The instruments used to evaluate achievement in high school physics should take account of the unique nature of physics and of the diverse nature of the human intellect. The advocacy, by such educational leaders as Bruner and Schwab, of inquiry approaches to learning, lead to the conclusion that we must teach, and presumably evaluate, physics as it really is. Bruner states that:

> Intellectual activity anywhere is the same, whether at the frontier of knowledge or in a third-grade classroom.[1]

This statement would seem to reinforce the contention that test items ought to recreate the reality of physics.

The problem might be most succinctly stated thus--what skills or abilities (as defined and measured by the Differential Aptitude Test battery) are most useful in correctly responding to five different kinds of test items in physics?

The research has been directed by an hypothesis which may be stated as follows:

The degree to which the spatial-mechanical factor operates as a predictor of success in physics tests depends on the kind of test items which are used. Test items in which the verbal content is limited by the use of diagrams will be more closely related to the Space Relations and Mechanical Reasoning subtest scores of the Differential Aptitude Test battery than will verbal, numerical, or graphic test items.

---

[1]J. S. Bruner, The Process of Education (New York: Vintage Books), 1960, p. 14.

A complete description of the test items and of the statistical treatment of the data will be given in Chapter III.

II. JUSTIFICATION OF THE STUDY

With the retention of ever larger numbers of students in the secondary schools, those who do not have highly developed verbal skills may be at a disadvantage when faced by the usual kind of test item. The need for high school graduates who are competent to take further technical training is urgent, and any factor which would limit the supply of such students should, if possible, be mitigated.

Civil and mechanical engineers, chemists, structural geologists, physicists, architects, and technologists in these fields, to name but a few, are required very frequently to perform mental manipulations of three dimensional objects. A high degree of spatial ability is therefore necessary for the adequate performance of their jobs. Confirmation of the foregoing statements may be found in the studies conducted by Dr. Anne Roe which are described in Chapter II.

Other studies have attempted to show the relationship between mechanical-spatial ability and physics achievement. Experience confirms that students who have the ability to manipulate three dimensional objects in their imagination have a better chance of solving physics problems. The relationship has not been as clear as might be expected however and the possibility exists that the fault lies with the kind of examination which has been used to measure physics achievement.

A recent study in England by Lewis has found a significant spatial factor in science tests of elementary school children. Lewis

has suggested:

> A further point is that of scientific attainment had been measured by school examinations (instead of by the objective tests which Lewis used), the variance from verbal and possibly other group factors would almost certainly have been greater. Thus, with essay type answers, variations in the accuracy and fluency of verbal expression would inevitably appear, and so augment the verbal factor. Again, if the test of examination required the use and interpretation of diagrams, a spatial factor might in this way be brought in.[2]

On the basis of recent experience in Manitoba and elsewhere, it would appear that most educational jurisdictions are eliminating the formal, centralized system of provincial examinations in favor of shorter and more frequent tests. It would seem desirable therefore that teachers design these tests as carefully as possible and that they know what kinds of skills and aptitudes are required to answer the test items. The tests that are given should be fair in sampling course content and should also be fair in sampling the differing skills and aptitudes which are needed to respond to the test items.

If one accepts the assumption, borne out by experience in teaching physics, that many of the concepts of physics involve the mental manipulation of three dimensional objects which is known as spatial relations or spatial visualization, then at least some of the test items used to evaluate achievement should require the use of spatial visualization. A statement by Dressel is appropriate to this

---

[2]D. G. Lewis, The Factorial Nature of Attainment in Elementary Science, Br. J. Ed. Psych., 1964, 34: 1-9.

argument:

> . . . since one of the axioms of measurement is that objectives not tested in examinations are not real objectives to students, it behooves every teacher to include items in examinations which measure accomplishment of all the real objectives of the course.[3]

It is possible to summarize the reasons for conducting this study rather simply. They are:

a. The need to know more about how test items work in relation to the kinds of skills and aptitudes that are used in responding to them.

b. The need to construct test items in physics which require the highly relevant aptitudes of mechanical reasoning and space relations.

c. The need to construct test items in physics which simulate in a practical and useable form, the observations which students make in the laboratory.

III. DEFINITIONS AND DELIMITATIONS

A. Definitions

Reference is made throughout this study to five different kinds of test items which are described as verbal, numerical, graphic, diagrammatic-verbal, and pure diagrammatic items. The distinctions

---

[3] P. Dressel, How the Individual Learns Science, Rethinking Science Education, Fifty-Ninth Yearbook of the National Society for the Study of Education (Chicago: University of Chicago Press), 1960, p. 59.

among these five kinds of items were established by simple and straightforward definitions:

1. Verbal items were designed so that both the stem of the question in which the question is posed and the distractors or the possible answers were in the form of words only.

2. Numerical items were designed so that the stem of the question was in the form of words and numbers requiring that a calculation be made in order to enable the student to choose correctly from among the five distractors which were in numerical form.

3. Graphic items were designed so that the stem of the question was in verbal form and the distractors in the form of different graphs from which the correct response was chosen.

4. Diagrammatic-verbal items were designed so that the stem of the question contained, as an essential part, a diagram as well as a verbal portion and the distractors were in verbal or numerical form.

5. Pure diagrammatic items were designed so that both the stem of the question and the distractors were in the form of diagrams. In all cases the verbal instructions of a few words were kept as uncomplicated as possible.

B. Delimitations

The study was conducted using as subjects students who had elected PSSC Physics 200 in a course which was designated as University Entrance. These students were already doing well in school as indicated by their selection of a series of courses which lead to University.

They had also elected a physics course which has the reputation of being difficult. It was to this group that the study was limited.

The findings of the study were limited to the subjects of optics and basic topics in physics as presented in the PSSC Physics course, second edition, Chapters 2 to 4 and Chapters 11 to 15. These topics in physics were felt to be particularly well suited to testing by items of a diagrammatic nature.

The accuracy and applicability of the study was obviously determined to a high degree on the adequacy with which the *Differential Aptitude Test* battery measured those skills and aptitudes which were used as predictors in the regression equation.

The applicability of the study to other classes of students depends upon the representativeness of the students used in the study. The students were unselected except as they had elected to take physics and as it was possible for the investigator to complete the testing program. A certain diminution of the sample was experienced because of the practical difficulties of administering all of the *Differential Aptitude Tests* to all of the students in the test group.

Three of the *Differential Aptitude Tests* were not administered to the test group because the aptitudes tested were not considered to be relevant to the study. The tests which were omitted were the *Clerical Speed and Accuracy* test, the *English Usage Spelling* test, and the *English Usage Grammar* test.

IV. ORGANIZATION OF THE THESIS

The following parts of this thesis will include, in Chapter II,

a summary of the related literature. In the second chapter, two kinds of related literature will be considered; an investigation of what certain authorities have said about the nature and structure of the human intellect, and a review of studies which attempt to use the assumed structure of the intellect to predict academic success in school science.

Chapter III will describe the nature of the investigation, the kind of test items used, the Differential Aptitude Test battery, the students who were subjects in the testing procedure, and the procedures which were followed in testing.

Chapter III will include a description of the methods which were employed to analyze the data. The technique, known as stepwise regression, was employed so that a numerical answer might be given to such questions as--what combination of verbal, numerical, abstract, mechanical, and spatial reasoning best predict student performance on physics verbal test items, numerical test items, graph test items, diagrammatic-verbal test items, and pure diagrammatic test items?

The results of the investigation will be presented in Chapter IV. The basis of acceptance or rejection of the hypotheses proposed in Chapter III will be given as well as a summary in verbal and tabulated form, of the test and analysis results.

The final chapter, Chapter V, will draw some conclusions which are warranted by the results of the statistical analysis. Recommendations will be made for certain practices in testing physics achievement. Finally, some further studies will be recommended to further the knowledge we have of students in secondary school science.

CHAPTER II

REVIEW OF THE LITERATURE

The specific topics which were researched in connection with this study were: the psychological background of the study including the currently accepted beliefs of psychologists about the nature of human intellectual abilities and four studies in which these abilities are investigated; and the relationship between these human intellectual abilities and school success in physics particularly and science in general.

I. THE STRUCTURE OF THE HUMAN INTELLECT

The conceptions of the structure of the human intellect generally accepted by American and British psychologists differ. The American concept suggests that the primary mental abilities (the term used by Thurstone)[4] consist of V (verbal), N (number), P (perceptual speed), M (rote memory), I (inductive reasoning), D (deductive reasoning), W (word fluency), S (space or visualization). The British conception, as put forth by Vernon,[5] assumes a general factor, g, so called by Spearman;[6]

---

[4] L. L. Thurstone and T. G. Thurstone, Factorial Studies of Intelligence (Chicago: University of Chicago Press), 1941, p. 4.

[5] Phillip E. Vernon, The Structure of Human Abilities (London: Methuen and Co. Ltd.), 1961, p. 73-74.

[6] C. Spearman, The Nature of Intelligence and the Principles of Cognition (London: MacMillan and Co. Ltd.), 1927, p. 5.

and two major group factors, v:ed (verbal-numerical, educational) and k:m (mechanical and spatial subfactors). The essential difference is that the Americans have attempted to explain differences in human intellect in terms of a number of specific mental abilities whereas the British have attempted to explain them in terms of a broad, general factor and two major groups of subfactors.

Vernon's point of view is the one to which reference will be made in this thesis. The nature of the human intellect, as proposed by Vernon is shown in Figure 1. This structure is representative of the British concept.

A somewhat different conception of the structure of the intellect is that proposed by Guilford.[7] As shown in Figure 2, Guilford proposes several different intellectual factors. Each of the factors is expressed in terms of three classes of content, namely, figural, symbolic, and semantic.

The figural content of each of the intellectual factors is visual, auditory, or kinesthetic in form. It is recorded by the senses and the visual content is contained, in this study, particularly in those physics test items which have diagrammatic content. Presumably a kind of visual memory factor may also be involved in the process of responding to the verbal and numerical items.

The symbolic content is expressed in numbers, letters or other symbols. In the context of this study the process of responding to numerical and graphic test items presumably depends to some degree upon

---

[7]J. P. Guilford, Personality (New York: McGraw-Hill Book Co., Inc.), 1959, pp. 360-365.

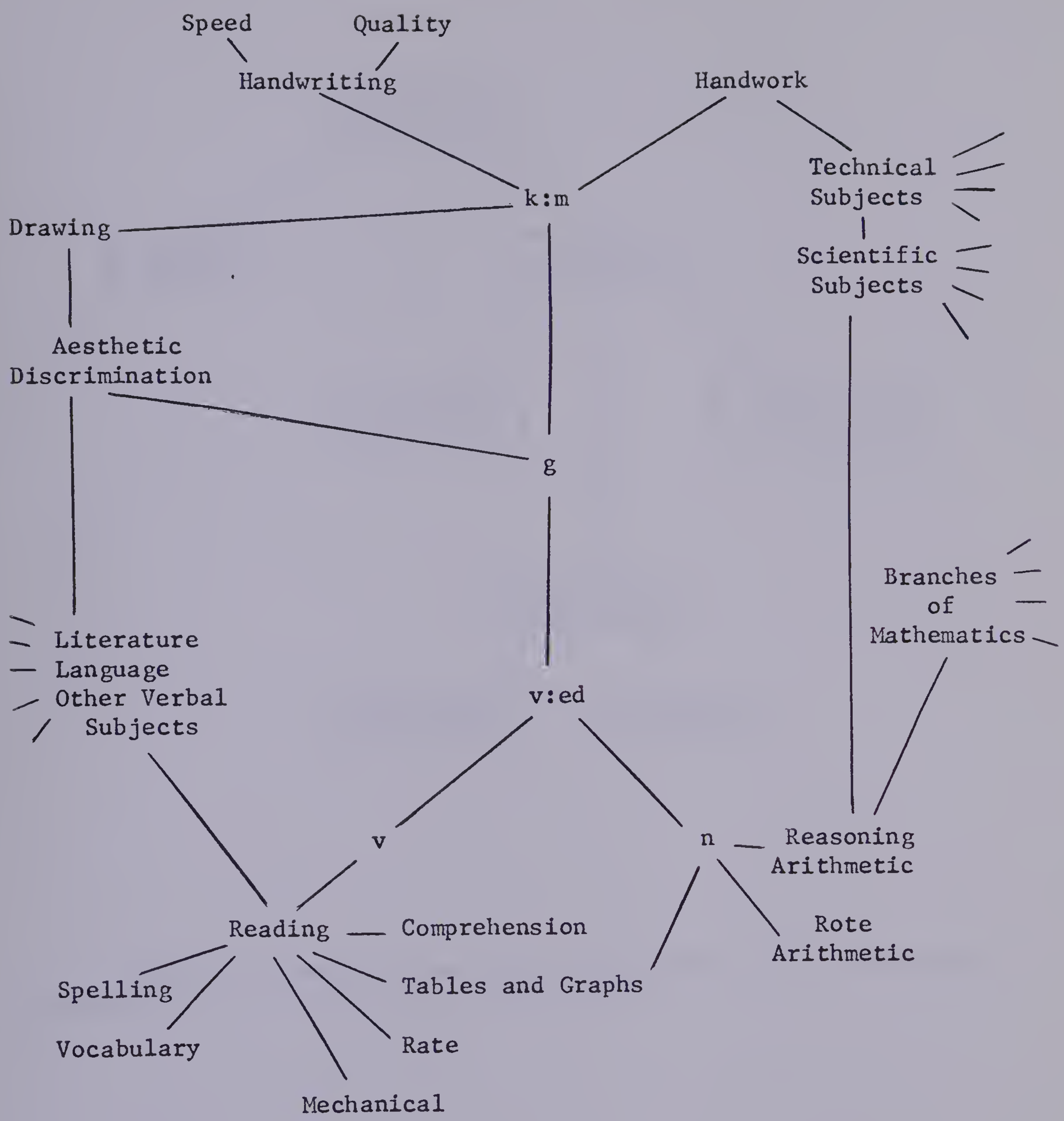


Figure 1. The Structure of Educational Abilities

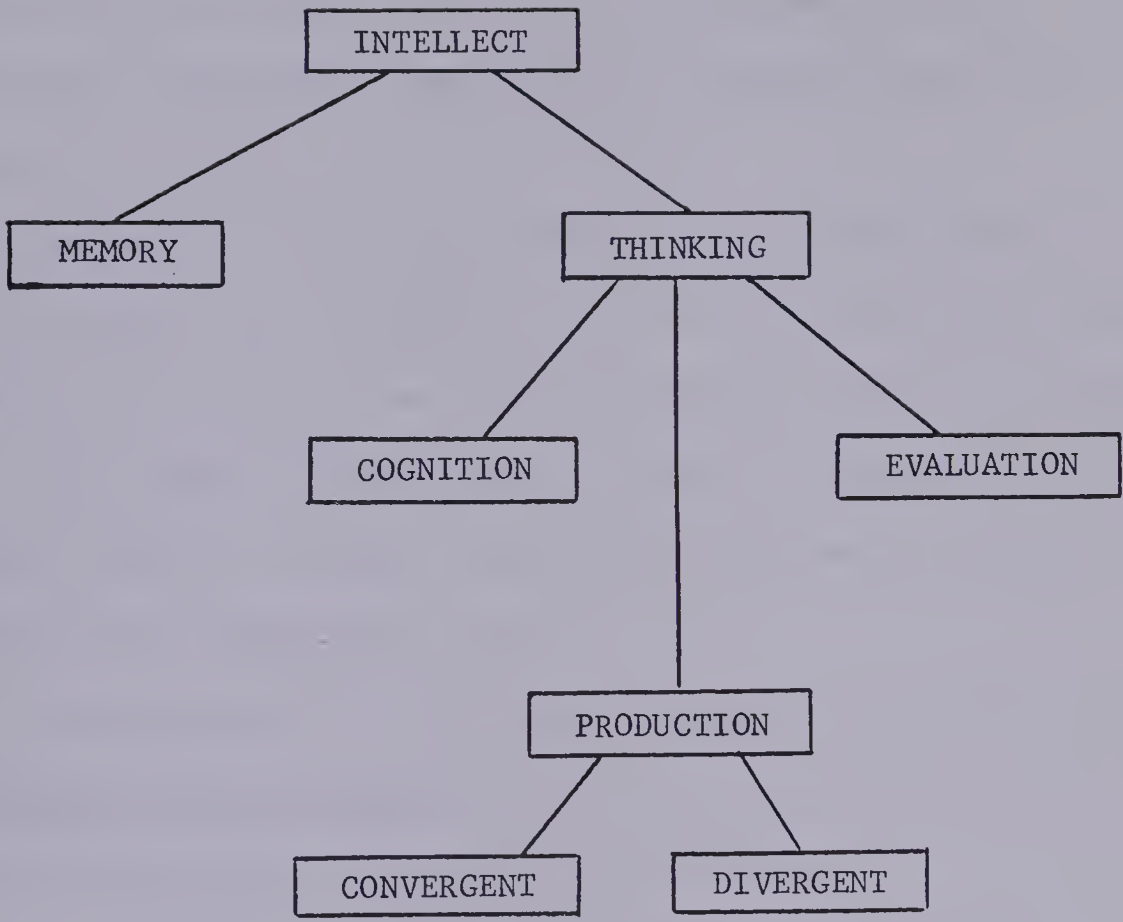


Figure 2. General plan of the relationship among the various main categories of intellectual abilities.[8]

[8] J. P. Guilford, <u>Personality</u> (New York: McGraw-Hill Book Co., Inc.), 1959, p. 360.

the symbolic content of these items.

The semantic content is expressed in the form of words and the abstract ideas expressed in word form. The semantic content of the physics test items may be important in the physics test items in which instruction is given in the form of words.

Whether one accepts the American view of many specific factors, or the British view of a general factor plus a more limited number of subfactors, the place of the spatial-mechanical factor is an important one. On the basis of either view, one might reasonably argue that tests of physics ought to contain items which would require students to use this ability in responding to them.

The introduction to the fourth edition to the Manual for the Differential Aptitude Tests puts the case for a battery of tests to measure "intelligence" in the following statement:

> Subsequent to the 1920's there was a growing recognition of the need for measures of many aspects of mental ability. The research and theories of Thorndike, Kelly, Spearman, Thomson, Thurstone and others have made us increasingly aware that so-called intelligence is not a unitary trait--it is composed of many abilities, which are present in different humans in varying amounts.[9]

Buell has outlined Piaget's five stages of cognitive development and summarizes them as:

i) Sensori-motor activities upon things.
ii) Pre-operational actions upon things.
iii) Intuitive operations with things.

---

[9] S. K. Bennett, H. G. Seashore, and A. G. Wesman, Differential Aptitude Tests, Fourth Edition Manual, Forms L and M (New York: The Psychological Corporation), 1966, p. I-1.

iv) Concrete operations with things.
v) Contemplation about things.[10]

Buell continues:

> It is the latter two (characteristic of post pubertal youth) with which secondary school teachers of science are concerned.[11]

This statement implies that, not only from the point of view of the nature of physics, but also because of the developmental stage of the majority of students, the laboratory ought to be an important part of a secondary school physics course. Evaluation of student achievement should, therefore, make use of test items which reproduce in a recognizable and practical way, the students' experiences in the laboratory. Because individual tests of laboratory skills are impractical and time consuming, it is reasonable to try to reproduce these laboratory situations in the form of diagrammatic test items.

In the fifth of Piaget's stages of cognitive development,[12] "contemplation about things," the argument for test items which involve diagrams and which call for visualization or mental manipulation of objects appears to be well founded. Piaget's fifth stage appears to involve the same kind of mental processes as those aptitudes tested by the _Abstract Reasoning_, and particularly the _Mechanical Reasoning_, and _Space Relations_ tests of the DAT battery.

A staunch advocate of the predictive value of spatial tests, I. M. Smith,[13] argues that British schools ought to make greater use of

---

[10] R. R. Buell, _Piagetian Theory into Inquiry Action, Sc. Ed._, Vol. 51, No. 1, 1967, p. 21.

[11] _Ibid._

[12] _Ibid._

[13] I. MacFarlane Smith, _Spatial Ability_ (London: University of London Press), 1964.

spatial tests as one criterion in selecting students for technical schools and even for grammar schools, if the student has elected a science course.

II. RELATED STUDIES

A. The Roe Studies

In a series of studies Anne Roe investigated the psychological nature of eminent scientists and also reported in her book, The Psychology of Occupations, on the psychological characteristics of successful men in a wide variety of professional fields.

Of interest here are her reports on the nature of eminent scientists. Doctor Roe studied a total of 59 scientists judged by their peers to be eminent in their research. They included 19 biologists, 7 experimental physicists, 11 theoretical physicists, 14 psychologists, and 8 anthropologists. In her study Roe administered a series of three special tests to the group of scientists. They were a verbal test, a spatial test and a mathematical test. The scientists all wrote the three tests except for the physicists who did not write the mathematics test because it was too easy for them. She found that the theoretical physicists had the highest average verbal score and also the highest average spatial score.[14]

In her general book entitled The Psychology of Occupations, Roe quotes O'Connor's list of the characteristics of engineering executives.

---

[14] Anne Roe, The Psychology of the Scientist, Science, 134: 456-459, August 18, 1961.

Characteristics common to technical engineers are given as: structural visualization and subjective personality. The spatial visualization factor agrees with test results on creative physicists, who are also high in this capacity.[15]

B. Studies of the Spatial Factor in Relation to School Success in Physics

1. The Sharo Study

The Sharo Study[16] was conducted in Stuyvesant High School, New York City. Sharo used the specially designed tests of spatial, mathematical and verbal ability constructed by Roe for her study of eminent scientists. He also used the results of the College Regents examinations in Physics, Chemistry, Mathematics and English; the College Entrance Examination Board tests of Scholastic Aptitude, Physics, Mathematics and the verbal test.

Sharo found the intercorrelation among the variables. As far as the predictive value of the spatial test for physics success was concerned, the correlation between the spatial test and the physics examination was lower than for the verbal test and the physics examination.

---

[15] Anne Roe, The Psychology of Occupations (New York: John Wiley and Sons Inc.), 1966, p. 200, citing J. O'Connor, Characteristics Common to Engineering Executives (Hoboken, N. J.: Human Engineering Laboratory), 1938.

[16] E. A. Sharo, "Physics, Mathematics and Visual Spatial Relations," Unpublished Doctoral Thesis, Columbia Teachers College, New York, 1962.

	CEEB Physics	Regents Physics	SAT Verbal	Spatial (Roe)
CEEB Physics	x	.69	.69	.55
Regents Physics	.69	x	.60	.31
SAT Verbal	.69	.60	x	.31
Spatial (Roe)	.55	.31	.31	x

For such highly respected and widely used examinations as the Regents and College Entrance Board Exams, obviously the verbal content is very high. It is interesting to note that the correlation of the spatial test is exactly the same with the verbal test of the SAT as it is with the Regents Physics examination.

2. The Hukins Study

In Alberta, Hukins' study, "A Factorial Investigation of Science Tests" found that the Grade IX and X examinations loaded most heavily on the verbal and reasoning factors which he identified by factor analysis. He found no significant loadings on the spatial factor but suggested:

> It seems that the use of diagrams upon which questions are based is one useful method of minimizing the verbal content.[17]

3. The Watson-Blade Study

In a remarkable study carried out at Cooper's Union Engineering school, it was found that the spatial ability of students in first year engineering can be improved markedly by exposure to courses in

---

[17] A. A. Hukins, "A Factorial Investigation of Science Tests," Unpublished Doctoral Thesis, University of Alberta, 1963, p. 127.

descriptive geometry and mechanical drawing. Watson and Blade state, in part, in their findings:

> The authors believe that the data justify the conclusion that much of the success of an examinee taking the College Board spatial relations test rests upon a level of experience so far from the optimum, at the time of College admission, that two freshmen courses can raise the level of a group that is already one-half sigma above the general population tested by the College Board, to an average performance that is one and one-half sigmas above.[18]

On the basis of this study they go on to predict that a spatial relations test given at the end of first year engineering would be an excellent predictor of success in engineering school. They say, further, that spatial visualization scores are better than mathematics scores as predictors of the ability of a student to graduate as an engineer.

Myers[19] found equally striking results at the United States Military Academy. He found a similar improvement in Spatial Relations scores and also a high correlation (.65) with grades in military topography.

4. The Berdie Study

In a study of 472 students at the University of Minnesota Institute of Technology, Berdie[20] found that the Spatial Relations

---

[18] M. F. Blade and W. S. Watson, Increase in Spatial Visualization Test Scores During Engineering Study, Psychological Monographs, Vol. LXIX, No. 12, 1955, p. 1.

[19] Charles T. Myers, A Note on a Spatial Relations Pre-test and Post-test, Ed. and Psych. Meas., 13: 596-600, 1953.

[20] R. F. Berdie, The Differential Aptitude Tests as Predictors of Engineering Training, J. Ed. Psych., 42: 114-123, 1951.

subtest of the DAT and the American Council on Educational English exam along with the Numerical Ability subtest of the DAT were the best predictors of success in first year engineering.

The Berdie study does not agree with that of Watson and Blade probably because the latter was a longitudinal study. Berdie might well have come to the same conclusions as Watson and Blade had he followed the class to graduation. Indeed, the Watson and Blade study discovered that the Spatial Relations test results improved so remarkably during the first year of engineering that it is not surprising that Berdie, using the Spatial test as a predictor for first year engineering only, did not find it useful.

5. Selection of Pupils for Different Types of Secondary School in England

Following World War II the English education system instituted a series of reforms including the now famous eleven plus examination series by which children were selected for various kinds of secondary education. A great deal of testing was done to identify those instruments which would best predict the success of students in various kinds of secondary schools.

The conclusions of a symposium on the selection of students were reported by J.J.B. Dempster. Their conclusions were that no single test could be reliably used to select students for secondary schools, but that a battery of three general intelligence tests along with parental wishes form the best combination of factors to predict success in grammar schools. They concluded that the spatial factor was <u>not</u>

useful in guiding the choice of students for secondary schools.[21]

I. M. Smith, in his book Spatial Ability, disagrees with the conclusions of the Dempster group, stating:

> It would appear from these findings that the eleven plus selection procedure in Aberdeen seems to operate moderately successfully from the point of view of its long-term validity for selecting potential arts and medical students, but it appears to give preference to pupils who are likely to be less successful in university science courses. Since the English tests [Reading Comprehension and English Composition] tend to correlate negatively with criteria of success for science courses and positively for arts courses, there are grounds for suspecting that pupils who might succeed in university science are being rejected by the grammar schools because they lack the linguistic abilities required for success in the eleven plus examination.[22]

Smith contends that a combination of the Primary Head Teachers' Assessment plus the score of a special English examination plus the score on the NFER Spatial Test gave the best prediction of all around success in grammar school. As a further argument Smith submits evidence and opinion of Yates and Pigeon that the spatial factor has not been considered a useful predictor of grammar school and university success and therefore:

> The appearance of a test of spatial ability in this battery is deserving of comment. Such tests are not usually considered as useful predictors of success in grammar schools. It is possible, however, that the

---

[21] J. J. B. Dempster, "Symposium on the Selection of Pupils for Different Types of Secondary Schools," Br. J. Educ. Psych., 1948, 18, 3, pp. 121-133.

[22] I. M. Smith, op. cit., pp. 28-29.

> abilities that are measured by this kind of test are related to subsequent success in some branches of mathematics and science. It would seem to be desirable for further research to be undertaken to investigate this point.[23]

6. Prediction of Success in Physics as Quoted by the Differential Aptitude Test Publishers

In the manual of the Differential Aptitude Tests several small studies have been reported, showing the correlation between success in Physics and the subtests of the DAT. The evidence is inconclusive at best and is reported below in order to illustrate the wide range of correlation coefficients which have been obtained.

Physics Courses			Validity Coefficients between DAT and Course Grades in Physics				
			VR	NA	AR	MR	SR
Grade XI	boys		.43	.45	.41	.48	.48
Grade XII	boys		.31	.40	.30	-.32	.18
Grade XII	boys		.25	.40	.19	-.17	.11
Grade XII	boys		-.03	.04	.04	.40	.08
Grade XII	boys		.13	.26	.10	.35	.07
Grade XI	girls		.17	.34	.21	.34	.43
Grade XI		Prediction of DAT for course grades taken up to 3½ years later	.59	.60	.38	.47	.15
College 1st year			-.05	.24	.34	.26	.05
College 2nd year			.19	.14	.33	.21	-.07

VR - verbal reasoning
NA - numerical ability
AR - abstract reasoning
MR - mechanical reasoning
SR - space relations[24]

---

[23] I. M. Smith, ibid., pp. 29-30, citing A. Yates and D. A. Pigeon, Admission to Grammar Schools (London: Newner), 1957.

[24] George K. Bennett, Harold G. Seashore, Alexander G. Wesman, Differential Aptitude Tests, Fourth Edition Manual, Forms L and M (New York: The Psychological Corporation), 1966, pp. 5-14 and 5-15.

The results shown on the preceding table are, at best, equivocal. The efficiency of the SR factor in predicting physics results whether at the secondary or college level is not particularly high.

III. SUMMARY

Existing theories about the nature of the human intellect and the nature of eminent scientists in particular show the importance of the spatial factor of intelligence. Consideration of the visual or spatial nature of many of the basic ideas in physics confirms the importance of the spatial factor for an understanding of these basic ideas.

The studies of Blade and Watson, and Myers, indicate that the use of materials which give practice in the use of spatial concepts will improve the spatial relation scores markedly. This is experimental evidence supporting the use of diagrams in teaching. If diagrams and other two dimensional representations of apparatus are used in instruction then they should also be used in the evaluative process.

Studies such as that of Sharo show a significant correlation between the physics examination and spatial tests although it is not so great as the correlation with verbal reasoning tests. The Berdie study also shows a very limited usefulness for the Spatial Relations test as a predictor of success in engineering studies. In Britain there is much controversy over the usefulness of spatial tests in predicting success in grammar school and university courses.

In general, studies which attempt to use spatial relations measures to predict success in physics and closely allied technological applications of physics are equivocal. Studies such as those of Hukins

and Lewis have called for the use of examination items based upon diagrams so that the results of such examinations might be more closely related to the spatial factor.

CHAPTER III

THE TESTS, STATISTICAL PROCEDURES AND HYPOTHESES

A description and discussion of the tests used in this study is presented in this chapter. The statistical procedures for determining the reliability coefficients for the tests and the intercorrelations among the test scores are outlined. The final part of the chapter contains a statement of the hypotheses which were tested.

In order to place the contents of the third chapter in perspective, a restatement in broad terms of the central idea of the study may be useful. This study was designed to examine the relationship between achievement in a series of physics tests in which the amount of verbal and diagrammatic content varied systematically and intellectual aptitudes as measured by selected subtests of the _Differential Aptitude Test_ battery.

I. THE TESTS

A. _The Differential Aptitude Tests (DAT)_

The DAT subtests were chosen for use as the predictor variables in the investigation mainly because of the wide acceptance of the DAT test battery in North America. One advantage of the DAT lies in that it was specifically designed for use with secondary school students. Another factor in the choice was the availability of considerable research information regarding its validity and reliability. The DAT was chosen because it includes in a single battery, tests of all of the aptitudes with which the investigator was concerned in this study.

Although most of the comment regarding the DAT approves its usefulness and the size of the standardizing population, some questions are raise concerning certain other characteristics of the battery. Keats,[25] f xample, points out that first order correlations among the subtests are not necessarily true indicators of the relative contributions of the tests to the information provided by the whole battery. Schultz[26] questions the lack of information about the item analysis techniques which were used in constructing the tests. Schultz also points out the undue amount of overlap among the tests leading to high correlations among the tests.

It should also be noted in relation to this study that the DAT battery was developed for use in testing students from Grade VIII to Grade XII. Tests which purport to measure aptitudes over such a wide age range may not be sensitive enough to differentiate among a rather highly selected group such as students who have chosen physics at the Grade XI level. Keats[27] raises this question when he points out that irregularities are apparent in the distribution of the percentile scores, particularly at the upper and lower ends of the distribution. The percentile scores may be somewhat unstable and possibly biased for norms based on little more than two thousand cases.

Despite the objections noted, the DAT is probably the most widely

---

[25] J. A. Keats, The Differential Aptitude Tests, The Sixth Mental Measurements Yearbook; Edited by Oscar K. Buros (Highland Park, Ill.: The Gryphon Press), 1965, p. 1004.

[26] Richard E. Schultz, ibid., pp. 1005-1006.

[27] J. A. Keats, ibid.

used aptitude battery on this Continent. Its wide acceptance is probably its biggest virtue in this study, and makes possible the consideration of the extensive data upon which reported reliability and validity coefficients are based.

The tests used in this study are: Verbal Reasoning (VR), Numerical Ability (NA), Abstract Reasoning (AR), Mechanical Reasoning (MR), and Space Relations (SR). Each of these tests will be described in turn.

1. Verbal Reasoning

The Verbal Reasoning test employs verbal analogies to measure the ability of a student to reason in a verbal context. The student must first understand the meaning of the words used, then be able to perceive the relationship among the words and finally choose the correct pair of words from among the five pairs offered as possible responses.

Reliability is the "sine qua non" of tests. The reliability coefficient is a measure of the internal consistency of a test. It indicates the degree to which all of the items of the test measure the same thing. Reliability is statistically defined as the ratio of the observed variance for the test to the true variance where observed variance is equal to the sum of the true variance and the error variance. If tests are not reliable, it is not possible to use results from them with any degree of confidence.

The reliability of VR has been reported in the form of two reliability coefficients for each grade level and for both boys and girls. For the grade level of the students in this study, the reliability coefficients reported in the DAT Manual are .93 (odd-even scores) and

.93 (sample drawn from the standardizing population) for girls, and .93 and .92 for the boys.[28]

Validity is the term used to describe the ability of a test to measure what it purports to measure. Validity is determined by correlation of the test scores with outside criteria. By such means it is hoped to establish what a test measures.

Cronbach[29] describes three different kinds of validation: predictive validation (and a variation termed concurrent validation) which is a measure of the usefulness of a test to predict future performance; content validation, which measures the validity of the test in sampling the content of what it purports to measure, and, construct validation, which translates the results of a test into psychological terms and measures the ability of the test to quantify the subject's characteristics in terms of a particular psychological attribute.

Validity is specific--there is no such thing as "the validity" for a test. A test may be a valid measure for one group under a certain set of conditions, but not for another group or even for the same group under a different set of conditions. It is, however, possible to speak of the validity of a test for predicting success in a particular subject. Such validity may in turn be specific to prediction of a particular test for a particular group of students. As Cronbach states:

---

[28]George K. Bennett, Harold G. Seashore, Alexander G. Wesman, Differential Aptitude Tests, Fourth Edition Manual, Forms L and M (New York: The Psychological Corporation), 1966, p. 6-2.

[29]Lee J. Cronbach, Essentials of Psychological Testing (New York: Harper and Row Publishers), p. 122.

> Different tests have different virtues; no one test in any field is 'the best' for all purposes. No test maker can put into his test all desirable qualities. A design feature that improves the test in one respect generally sacrifices some other desirable quality.[30]

In describing the validity of the Verbal Reasoning test, the authors of the DAT battery state:

> The Verbal Reasoning and Grammar tests have relatively high validities for predicting grades in most courses. This may be accounted for, in part, by the academic custom of awarding grades in most subjects on the basis of written reports, essay questions, and other verbal responses. Consequently, ability to verbalize with conventional grammatical structure can influence grades even in science.[31]

Because validity is highly specific, the predictive validity coefficients between VR and course grades in physics, based on six different studies, were found to vary from a high value of .43 to a low of -.03 with a mean value of .21.[32] The very large range of validity coefficients of VR for course grades in physics suggests the advisability of attempting to determine why different tests yield such widely differing results. It is the purpose of this investigation to contribute some knowledge of the reasons for this variation.

Validity coefficients for VR with certain standardized achievement tests are also reported by the authors of the DAT. The most highly correlated tests with VR were reported to be the Iowa Tests of Educational Development (ITED), Correctness and Appropriateness of Expression, .87

---

[30] Ibid., p. 115.

[31] George K. Bennett, Harold G. Seashore, Alexander G. Wesman, Differential Aptitude Tests, Fourth Edition Manual, Forms L and M (New York: The Psychological Corporation), 1961, pp. 5-3 and 5-4.

[32] Ibid., pp. 5-14 and 5-15.

(boys) and .88 (girls); ITED, Use of Sources of Information, .87 (boys), .85 (girls).[33]

These coefficients tend to convey that the VR test does in fact test verbal skills and is thus most highly correlated to such things as abilities in English composition, word knowledge, and how to find verbal information. The skills involved in spelling, arithmetic computation, and study skills in social studies have been found not to be as closely related to VR as the above.

2. Numerical Ability (NA)

The Numerical Ability subtest of the DAT consists of items which measure the student's ability to perform basic arithmetic computations, including addition, subtraction, multiplication, division, square roots, cube roots, fractions, percent, ratio and proportion, and price discounts. The verbal content of the items is kept at a minimum so that the vocabulary of a student need only include meanings for terms such as add, subtract, multiply, divide, square root, and list price.

Reliability coefficients are reported for NA on the same basis as for VR, that is, for both boys and girls at the VIII to XII grade levels. The reliability coefficients reported in the DAT Manual for Grade XI girls are .92 (odd-even scores) and .91 (sample from the standardizing population) while for the boys the corresponding coefficients are .93 and .92.[34]

---

[33] *Ibid.*, pp. 5-34 to 5-38.

[34] *Ibid.*, p. 6-2.

The validity of NA subtest for the prediction of physics course grades was determined in six studies. The range of validity coefficients reported was between .04 and .45 with a mean value of .38.[35]

Validity coefficients are also reported for studies involving several standardized achievement tests. The tests with which NA was most highly correlated were the Stanford Achievement Tests Arithmetic Reasoning, .90 (boys), .86 (girls), and Arithmetic Computation, .89 (boys), .88 (girls); and the Iowa Tests of Educational Development Ability to do Quantitative Thinking, .86 (boys), .76 (girls).[36]

3. Abstract Reasoning (AR)

The Abstract Reasoning test is designed as a measure of a student's ability to reason with, and to recognize relationships among, abstract forms or diagrams which are not intended to represent objects. The test takes the form of a series of four diagrams. The student is asked to choose from five possible diagrams the one which would be next in the series.

For the Grade XI girls the reported reliability coefficients are .93 (odd-even scores) and .92 (sample taken from the standardizing population) and for Grade XI boys the corresponding coefficients are .92 and .90.[37]

---

[35] Ibid., pp. 5-14 and 5-15.

[36] Ibid., pp. 5-34 to 5-38.

[37] Ibid., p. 6-2.

The results of six studies establishing the validity for the prediction of course grades in physics are reported in the DAT Manual. The validity coefficients range from a low of .04 to a high of .41 with a mean of .21.[38]

The highest validity coefficients for AR with standardized achievement tests are with ITED Ability to do Quantitative Thinking, .76 (girls), .75 (boys); Stanford Achievement Tests Average Arithmetic, .80 (girls), .72 (boys); and Stanford Achievement Tests Arithmetic Reasoning, .78 (girls), and Arithmetic Computation, .79 (girls).[39]

In general, AR appears to have the highest validity for predicting scores on such things as Ability to do Quantitative Thinking, and Arithmetic Reasoning. While AR is generally more useful than the other DAT subtests for predicting non-verbal achievement, there are particular achievement tests of a highly verbal nature with which AR is highly correlated.

4. Mechanical Reasoning (MR)

The Mechanical Reasoning subtest of the DAT measures the ability of a student to perceive the manner in which certain mechanical devices operate. The items used to measure this ability are given in the form of diagrams which represent actual mechanical devices. The verbal part of the item is simply worded and the response of the student involves a choice among three alternatives. It is claimed by the authors of the DAT that formal training in physics does not significantly

---

[38] Ibid., pp. 5-14 and 5-15.

[39] Ibid., pp. 5-34 to 5-38.

change MR scores.

The reliabilities reported for the MR subtest are not as high as those of the other tests. For girls in Grade XI the reliability coefficients are .83 (odd-even scores) and .83 (sample from the standardizing population). For boys the corresponding coefficients are .90 and .89.[40]

Six studies were conducted to establish the validity of MR for the prediction of physics courses. The range of validity coefficients reported is from -.32 to .48 with a mean of .18.[41] The reason for the very low validity coefficient is that two of the six studies yielded negative validity coefficients.

In studies involving standardized achievement tests, the validity coefficients fail to give a clear picture. The three highest coefficients for boys were with STEP *Science* (.67), and *Math* (.65), and the ITED *Ability to do Quantitative Thinking* (.63). For girls, however, the three highest coefficients were MR with STEP *Social Studies* (.63), ITED *General Background in Natural Sciences* (.62), and ITED *Interpretation-Literature* (.60).[42]

Perhaps the most one may say about the validity of MR test is that it appears to predict non-verbal achievement adequately in the majority of cases. The validity of this test, perhaps more than for the others, appears to be specific to the group being tested and/or to the standardized test with which it is being correlated. The lowest validity

---

[40] *Ibid.*, p. 6-2.

[41] *Ibid.*, pp. 5-34 to 5-38.

[42] *Ibid.*

coefficients were observed for MR with the Metropolitan Achievement Tests (Advanced Battery, Form AM), for both girls and boys.

5. Space Relations (SR)

The Space Relations test of the DAT measures the ability to mentally manipulate three-dimensional objects. It involves the process of visualizing an object in other positions. Terms such as space visualization have been used to describe closely related abilities.

The Space Relations test consists of a two-dimensional pattern which, when folded, would produce a three-dimensional object. The test displays four alternative responses in the form of diagrams of the three-dimensional object, one of the four forms representing the correct object.

The SR test has been chosen for inclusion in the battery used in this study because the ability to visualize an object as it moves or turns is often necessary in physics. For example, to predict the direction in which a ray of light will be refracted as it enters a new medium obliquely requires a degree of spatial ability. To imagine the position of the shadow cast by an object in the presence of a light source also involves spatial ability. A degree of spatial ability is frequently necessary to imagine the relative positions of objects described in the verbal portion of a problem.

The reliability coefficients reported in the DAT Manual for the SR test were, for Grade XI girls, .95 (odd-even items), and .95 (sample drawn from the standardizing population). For the boys, the

corresponding coefficients were .95 and .95.[43]

The results of six studies dealing with validity of SR for the prediction of course grades in physics are reported in the DAT Manual. The studies report validity coefficients between SR and physics ranging from .07 to .48 with a mean value of .23.[44]

The validity of SR for the prediction of scores on several standardized achievement test batteries gives some indication of the kind of academic achievement most closely related to SR. For girls the three highest coefficients were found to be with STEP _Mathematics_ (.77), _Stanford Achievement Tests Arithmetic Computation_ (.70), and ITED _Use of Sources of Information_ (.68). For boys the three highest validity coefficients were between SR and ITED _Ability to do Quantitative Thinking_ (.71), STEP _Science_ (.66), and STEP _Mathematics_ (.65).[45]

Although the results quoted could hardly be said to be conclusive, the higher correlations tend to be with the less verbal subject matter areas.

The _Differential Aptitude Tests_ were administered in accordance with the instructions of the authors during the spring term of 1970. The majority of the tests were machine-scored by the Guidance Services Branch of the Manitoba Department of Youth and Education. Single tests for four of the students in the sample were carefully hand-scored by the investigator using the key provided by the publishers of the DAT.

---

[43] _Ibid._, p. 6-2.

[44] _Ibid._, pp. 5-14, 5-15.

[45] _Ibid._, pp. 5-34 to 5-38.

B. <u>The Physics Tests</u>

The physics test items were sixty-five in number. One item was discarded because of an error in one of the diagrams of the test. The entire test was used as the pre-Easter examination for the PSSC Physics 200 course in Grade XI at Vincent Massey Collegiate, Fort Garry, Manitoba during the last week of March 1970. The test items were all designed as multiple choice items with five alternative answers.

The multiple choice form for the test items was used because of the objectivity of scoring such items and because of the increasing frequency with which this type of item form is used in secondary school evaluation. Finally, and most importantly, the multiple choice type of test enables the investigator to use well-accepted standard procedures for estimating the reliabilities of the tests.

All responses were recorded by the students on IBM 1230 answer sheets and the sheets were scored by the Data Processing Centre at the University of Manitoba.

A first draft of the tests was administered to the Grade XI class at Vincent Massey as a physics test in March of 1969. On the basis of this trial, alterations were made before its use the following year.

The use of the test as a part of the regular testing program, in fact, the only extensive examination written by the Grade XI physics students in 1970, was thought to make it an appropriate measure of physics achievement in two ways. First, it was written by all of the students at the same time under careful supervision. Second, the students were more likely to be well motivated than would have been the

case under less formal conditions.

The portion of the PSSC Physics course tested included Chapters 2, 3, and 4 of Part I and Chapters 11, 12, 13, 14 and 15 of Part II from the 1965 revision of the PSSC Physics text.

The usual types of multiple choice items were examined and divided into categories which were distinguishable on the basis of the amount of verbal and diagrammatic content. The aim in making the division of the items into the various categories was that the basis for categorization be obvious and that the categories be mutually exclusive. The divisions would therefore consist of specific types of items which, it was hoped, would have specific characteristics when analyzed statistically.

The five divisions into which the items were placed are: Physics-verbal (PhV), Physics-numerical (PhN), Physics-graphs (PhG), Physics-diagrammatic verbal (PhDV), and Physics-pure diagrammatic (PhPD). Of the five categories, only the pure diagrammatic items are not commonly found in physics tests. The tests are presented in Appendix C. Each will now be described in turn.

1. Physics-verbal (PhV)

The verbal test items in physics were constructed with a wholly verbal stem and verbal distractors. Five alternative responses were provided and the selection of the correct response did not require any numerical calculation. There were 15 verbal items in the whole test. In order to minimize the fatigue factor and to apportion any unfinished

items more or less equally among the five subdivisions of the test, the verbal items were scattered throughout the test in an irregular pattern. The time factor is unlikely to be importance since three hours was permitted for answering the test and all of the students completed it.

Item number eight is presented as an example of the verbal item type:

> 8. In 'stopping' certain kinds of motion which are too fast to be seen with the unaided eye we may use a stroboscope. The stroboscope is useful for stopping only one of the following. Which one?
> a) A rifle bullet.
> b) A single wave travelling along a stretched string.
> c) An airplane propellor.
> d) A light ray.
> e) The shock wave in front of a supersonic airplane.

2. Physics-numerical (PhN)

The numerical test items were constructed with a verbal stem and numerical distractors. Five alternative responses were given, only one of which was correct. All of the numerical items required some calculation in order to arrive at the correct response. The test contained twelve items which were distributed in an irregular pattern throughout the whole test. The irregular placement of the items was again to minimize fatigue or time factors.

Number fifty-three is presented as an example of the numerical item type:

> 53. A pencil of light passes from water ($n = 1.33$) into a transparent plastic block. The angle of incidence is 48° and the angle of refraction is 36°. The relative index of refraction from water

to plastic is: (sin 48^{o} = .743, sin 36^{o} = .588).
a) 1.26 b) 1.78 c) 1.33
d) 1.68 e) 1.06

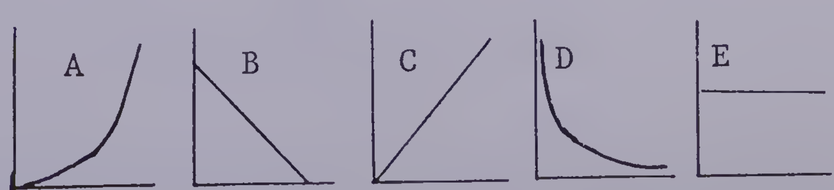
3. Physics-graphs (PhG)

The graphical items were constructed with a verbal and/or numerical stem and with distractors which were all in the form of graphs. The correct response was obtained by means of numerical calculations and a selection of the graph which represented the correct mathematical relationship between the variables. Five alternative graphs were given from which the correct response was selected by the student.

Ten such items were irregularly distributed throughout the test. to minimize any fatigue or time factor. Because two or three items referred to the same group of graphs, these items were grouped together in two or three item blocks. The blocks were irregularly spaced throughout the test.

As an example of the graph type of item, number thirty-three is displayed below:

Questions 33 to 35 relate to an optical experiment. Each question is to be answered with one of the five graphs given below. The horizontal axis represents the object distance. A graph may be used once, not at all, or more than once.



An object is placed at the principal focus of a converging lens and moved slowly away from the lens.

33. Which graph represents the location of the image as the object moves away?

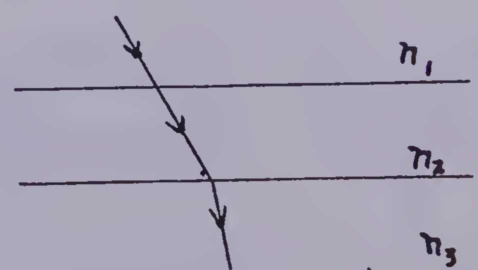
4. Physics-diagrammatic verbal (PhDV)

The diagrammatic-verbal items were perhaps the most variable item type in the test. In some instances a numerical calculation was required before the correct response could be selected from the five alternatives. In other items a mental spatial manipulation of the diagram was necessary in order to make the correct selection. In all items the verbal content was substantial and a number of the instructions were quite complex. In every item a diagram was included as a part of the stem of the item so that reference had to be made to the diagram in the solution of the item.

There were eighteen such items in the test, one of which was subsequently deleted from the data. The items were irregularly placed throughout the test for the reasons given previously in relation to the verbal, numerical and graph items.

Item number thirty-six is presented as an example of the type of item which was included in this category:

The diagrams below show the path of a narrow pencil of light as it passes through three different transparent materials. Select the combination of indices of refraction which would account for each diagram.



a) $n_1 > n_2 > n_3$ b) $n_1 > n_2 = n_3$

c) $n_1 = n_2 > n_3$ d) $n_1 = n_2 < n_3$

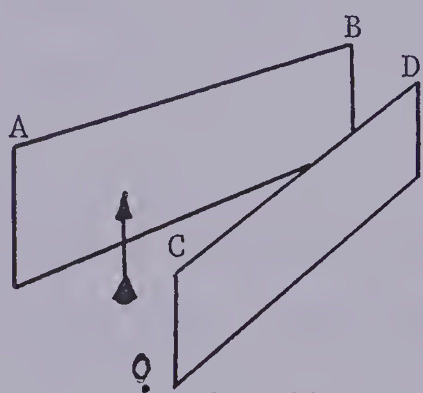
e) $n_1 = n_2 = n_3$

5. Physics-pure diagrammatic (PhPD)

The pure diagrammatic type of item was constructed so that both the stem and the distractors were in the form of diagrams. The verbal part consisted of very simple instructions. The vocabulary used was kept as simple as possible consistent with clarity and accuracy. In each item of this category, five alternative responses were again presented.

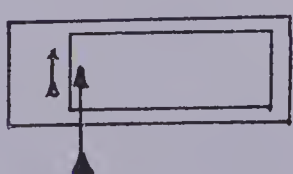
There were ten pure diagrammatic items included in the test. Item number sixteen is presented as typical of the items in this category:

16. A B and C D are plane mirrors, the silvered sides face one another.

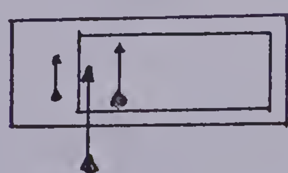


The observer at O will see:

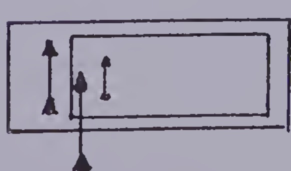
a)



d)

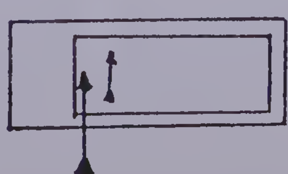


b)



e)

c)



II. STATISTICAL PROCEDURES

A number of statistical tests and procedures were used to analyze the results from the physics testing program and to answer questions about the relationship between the scores on the physics tests and the scores on the Differential Aptitude Tests.

The first task was that of establishing the reliability of the physics tests. The means and standard deviations were obtained for each subtest. The test responses were subjected to an item analysis. Separate analyses were performed for each subtest because the particular reliability coefficient calculated is an estimate of the internal consistency of the test, and the test was specifically constructed to measure more than one factor. The frequency distribution of scores, the individual item difficulties and the biserial correlation coefficients were also obtained.

The Kuder-Richardson formula 20 was used to obtain an estimate of the reliability of each test. The computational formula is:[46]

$$r = \frac{n}{n-1} \quad \frac{S^2 - \Sigma pq}{S^2}$$

where
- r = the reliability coefficient of the test.
- n = the number of items in the test.
- S = the standard deviation of the distribution of scores on the test.
- p = the proportion of students who responded correctly to the item.
- q = 1 - p

---

[46] Robert W. B. Jackson, George A. Ferguson, Studies in the Reliability of Tests, Bulletin Number 12, Department of Educational Research (Toronto: University of Toronto Press), 1941, p. 31.

Reliability estimates for the subtests of the DAT were not computed because extensive reliability data for the DAT is available in the manual published by the authors.

The emphasis in this study is on the relationship between the aptitudes tested and performance on specific types of physics items. Correlations among the physics tests were computed in order to determine the extent of common variance. The intercorrelations among the DAT subtests were required in order to permit inferences concerning the independence of the aptitude measures.

The intercorrelations between the DAT and the physics tests were of most interest to the investigator because these would indicate the relationship between the various types of achievement in physics and the five aptitudes measured by the DAT.

Stepwise Regression Analysis

The investigation of the relationship between the physics tests and the Differential Aptitude Tests was carried out by means of a variation of multiple regression analysis called "Stepwise Regression Analysis." The object of this statistical procedure is to determine the best linear regression model of the form

$$Y_1 = b_o + b_1X_1 + b_2X_2 + \dots b_5X_5 + e \qquad (1)$$

for each of the criterion variables (physics tests). In the model above Y_1 = the scores of the physics tests (the dependent or criterion variable); $X_1, X_2, X_3, \dots$ = the scores of the DAT (the independent or predictor variables); $b_0, b_1, b_2, \dots$ = the estimates of the coefficients which make the sum of the e^2 a minimum, and e is the error

term (the difference between the predicted and actual values of the dependent or criterion variable).

Equation (1) is approached in the stepwise procedure through a series of intermediate regression equations of the form

$$Y = b_0^{1} + b_1^{1} X_1 + e^{1}$$
$$Y = b_0^{11} + b_1^{11} X_1 + b_2^{11} X_2 + e^{11}$$
$$Y = b_0^{111} + b_1^{111} X_1 + b_2^{111} X_2 + b_3^{111} X_3 + e^{111}$$

In every case the first variable (X_1) to be entered into the regression equation is the one which is most highly correlated with the criterion. The coefficients b_1, b_2, . . . are determined from the multiple regression analysis. Where only two variables are entered into the regression equation (the first step shown above), the coefficient b_1 may be determined from:

$$b_1 = \frac{S_Y}{S_X} r_{XY}$$

Where S_Y and S_X are the standard deviations for the two tests and r_{XY} is the correlation between the two variables.[47]

In the more complex equations containing more than one predictor variable, the coefficients b_2, b_3 etc. are related to the partial correlation coefficient between Y and X_2, X_3, . . . , with the effects of the intercorrelation between Y and all preceding variables removed.

In order to determine whether the addition of a variable to the

---

[47] N. R. Draper, H. Smith, Applied Regression Analysis (New York: John Wiley and Sons Inc.), 1968, p. 35.

stepwise regression model makes a significant contribution to the prediction equation, the F ratio is computed. The F ratio may be represented as:

$$F = \frac{MS_R}{S^2}$$

where MS_R is the mean square due to regression and

S^2 is the mean square due to residual variation.[48]

The F ratio is distributed with n - 1 and N - n + 1 degrees of freedom where n is the number of variables entered into the regression equation and N is the number of observations (in this study N = 74).

In all multiple regression techniques, the problem becomes one of selecting which variable should be the next to be entered into the regression equation. The selection, in Stepwise Regression Analysis, is made by using the variable whose partial correlation coefficient with the criterion is largest. At each "step" the significance of the contribution made by the most recent variable to be entered into the regression equation is tested by the F test.

This technique is particularly applicable to the present study because of the high intercorrelation among the predictor variables. These intercorrelations indicate that there is considerable common variance among the predictor variables. The process of estimating the partial correlation coefficients is employed because it is necessary to remove the effect of the common variance shared with the previous predictor variables before attempting to analyze the variance of a

---

[48] Ibid., pp. 24-25.

newly added predictor variable.

III. HYPOTHESES

The hypotheses to be tested by the correlational analysis were:

A. There are no significant ($p<0.05$) zero-order correlations between the scores on verbal (PhV) test items in Grade XI PSSC Physics and (1) the *Verbal Reasoning* scores of the *DAT*; (2) the *Numerical Ability* scores of the *DAT*; (3) the *Abstract Reasoning* scores of the *DAT*; (4) the *Mechanical Reasoning* scores of the *DAT*; (5) the *Space Relations* scores of the *DAT*.

B. There are no significant ($p<0.05$) zero-order correlations between the scores on numerical (PhN) test items in Grade XI PSSC Physics and (1) the *Verbal Reasoning* scores of the *DAT*; (2) the *Numerical Ability* scores of the *DAT*; (3) the *Abstract Reasoning* scores of the *DAT*; (4) the *Mechanical Reasoning* scores of the *DAT*; (5) the *Space Relations* scores of the *DAT*.

C. There are no significant ($p<0.05$) zero-order correlations between the scores on graph (PhG) test items in Grade XI PSSC Physics and (1) the *Verbal Reasoning* scores of the *DAT*; (2) the *Numerical Ability* scores of the *DAT*; (3) the *Abstract Reasoning* scores of the *DAT*; (4) the *Mechanical Reasoning* scores of the *DAT*; (5) the *Space Relations* scores of the *DAT*.

D. There are no significant ($p<0.05$) zero-order correlations between the diagrammatic-verbal (PhDV) test items in Grade XI PSSC Physics and (1) the *Verbal Reasoning* scores of the *DAT*; (2) the *Numerical Ability* scores of the *DAT*; (3) the *Abstract Reasoning* scores of

the DAT; (4) the Mechanical Reasoning scores of the DAT; (5) the Space Relations scores of the DAT.

E. There are no significant (p/0.05) zero-order correlations between the pure diagrammatic (PhPD) test items in Grade XI PSSC Physics and (1) the Verbal Reasoning scores of the DAT; (2) the Numerical Ability scores of the DAT: (3) the Abstract Reasoning scores of the DAT; (4) the Mechanical Reasoning scores of the DAT; (5) the Space Relations scores of the DAT.

The hypotheses tested by the stepwise regression analysis were:

F. The efficiency of prediction of scores on verbal test items in Grade XI PSSC Physics is not significantly improved by adding to the Verbal Reasoning test scores of the DAT measures of Numerical Ability, Mechanical Reasoning, Space Relations, and Abstract Reasoning.

G. The efficiency of prediction of scores on numerical test items in Grade XI PSSC Physics is not significantly improved by adding to the Verbal Reasoning scores of the DAT measures of Space Relations, Numerical Ability, Mechanical Reasoning, and Abstract Reasoning.

H. The efficiency of prediction of scores on graph test items in Grade XI PSSC Physics is not significantly improved by adding to the Verbal Reasoning scores of the DAT measures of Mechanical Reasoning, Numerical Ability, Abstract Reasoning, and Space Relations.

I. The efficiency of prediction of scores on diagrammatic-verbal test items in Grade XI PSSC Physics is not significantly improved by adding to the Verbal Reasoning scores of the DAT measures of Space Relations, Mechanical Reasoning, Numerical Ability, and Abstract Reasoning.

J. The efficiency of prediction of scores on pure diagram test items in Grade XI PSSC Physics is not significantly improved by adding to the Mechanical Reasoning scores of the DAT measures of Verbal Reasoning, Space Relations, Abstract Reasoning, and Numerical Ability.

The sex variable was not investigated because of three factors. The number of girls in the sample was small (N = 20) compared to the number of boys (N = 54). The girls were highly selected because only the most able girls elect physics. Sex was not considered to be a significant factor in this study because the similarities between the boys and girls in the sample were considered to be greater than the differences between them.

CHAPTER IV

RESULTS AND DISCUSSION

The results of the investigation are presented and discussed in this chapter. The chapter is organized into four principal sections. The first is a presentation and discussion of the DAT, the scores as they reveal certain characteristics of the students in the sample, and the intercorrelations among the DAT subtests. The second section deals with the physics tests. The item analysis for the tests, a discussion of reliability in general, and the reliability coefficients for the physics tests are presented. Intercorrelations among the physics tests are also presented and discussed. The third section deals with the intercorrelations between the physics tests and the DAT subtests in relation to hypotheses A, B, C, D, and E. Finally the fourth section deals with the more detailed analysis of the intercorrelations between physics and the DAT as revealed by the Stepwise Regression Analysis.

I. ANALYSIS OF THE RESULTS FOR THE DIFFERENTIAL APTITUDE TESTS

Table I presents a summary of the data pertaining to certain characteristics of the students in the sample. The mean raw scores of the girls and of the boys for each of the five subtests of the DAT are presented. The corresponding percentile scores as they are given by the authors of the DAT are also presented. The separation of the scores of the girls from those of the boys was made because the percentile norms are given separately for girls and boys.

TABLE I

PERCENTILE SCORES FOR THE D A T--
DIFFERENTIATED BY SEX

	Test	Mean Raw Score	Percentile Score
Girls (N = 20)	Verbal Reasoning	40	80
	Numerical Ability	32	85
	Abstract Reasoning	41	75
	Mechanical Reasoning	47	80
	Space Relations	35	65
Boys (N = 54)	Verbal Reasoning	38	75
	Numerical Ability	31	70
	Abstract Reasoning	41	80
	Mechanical Reasoning	54	65
	Space Relations	42	65

The standardizing population for the DAT consists of more than 50,000 students drawn from 195 schools in 95 different communities in 43 states of the United States. These communities represent all of the major geographic areas, and within the geographic areas, efforts were made to ensure that the sample accurately represented the various sizes of community.[49] Thus the standardizing population may be said to approximate fairly accurately the high school population of the United States.

The particular group to which the sample of students from this study are compared numbers 2,250 boys and 2,300 girls.

In every case the mean performance of students from the sample was superior to that of the standardizing population. The superiority was most evident for the subtests, Verbal Reasoning (VR), Numerical Ability (NA), and Abstract Reasoning(AR). A good predictor of academic success, according to the authors of the DAT, is the VR + NA combination.[50] For these particular tests the girls' percentile scores were at the eightieth and eighty-fifth percentiles, whereas for the boys, the scores were at the seventy-fifth and seventieth percentiles. It may be that students who elect the PSSC Physics option are selected because of their superior verbal and numerical reasoning abilities.

It was noted that, in terms of absolute raw scores, the differences between the performance of the girls and boys in the sample is

---

[49]George K. Bennett, Harold G. Seashore, Alexander G. Wesman, op. cit., p. 3-2.

[50]Ibid., p. 5-3.

not large. This may justify, to some extent, the decision to treat the scores for both girls and boys together. The rather marked differences in percentile ranks are due to the generally lower scores for girls in the standardizing population on MR and SR. For MR the mean score of 47 corresponds to a percentile score of 80 for girls, while the mean score of 54 corresponds only to a percentile score of 65 for the boys. For NA the mean scores for girls and boys were 32 and 31 (very nearly the same), although the corresponding percentile scores were 85 for the girls but only 70 for the boys. This is again due to the lower NA scores for the girls in the standardizing population.

Table I is useful in identifying the kind of students in the sample. These students are academically superior, particularly in the verbal, numerical, and abstract reasoning skills to the standardizing population for the DAT norms.

Some reasons may be suggested for the demonstrated academic superiority of the sample. The students have probably been selected by performance on instruments which were largely verbal and numerical in nature. In elementary and junior high schools, the tests and examinations used to evaluate the achievement of students are generally heavily loaded with verbal and numerical items.

The intercorrelations among the DAT subtests are presented in Table II. It is important to note that all of the correlation coefficients between the subtests of the DAT are significantly different from zero at the five per cent level.[51] The tests are, therefore, not

---

[51]George A. Ferguson, Statistical Analysis in Psychology and Education (New York: McGraw-Hill Book Co. Inc.), 1959, p. 315.

TABLE II

THE MEANS, STANDARD DEVIATIONS, AND INTERCORRELATIONS FOR THE D A T SUBTESTS

(N = 74)

Test	VR	NA	AR	MR	SR
VR					
NA	.44*(19)				
AR	.34*(12)	.34*(11)			
MR	.48*(23)	.28*(8)	.37*(14)		
SR	.36*(13)	.32*(10)	.52*(27)	.69*(47)	
Mean	38.9	31.7	41.2	51.7	39.8
Standard Deviation	5.6	5.2	4.2	6.7	9.6

Percent of common variance in parentheses following correlation coefficients.

*Values which are significant at the 0.05 level.

independent. At least some of the aptitudes which are measured by the five subtests are common. Psychologists have used the term "g", the general factor, to explain this common variance. Spearman[52] has suggested the term "g" for this general factor and the practice has been continued, particularly among the British factor analysts.

The common variance for two tests which are correlated may be determined by computing the square of the zero-order correlation coefficient (r^2). The fraction of the variance which is common to the two tests is usually expressed as a percentage. The numbers in parentheses in Table II resulted from the calculation of r^2 multiplied by one hundred.

The intercorrelations, and hence the common variance, among the DAT subtests are quite large. The tests are not independent. A significant amount of what the test measures is measured by more than one test. When the scores of the DAT subtests are used as predictor variables in a multiple regression equation, as they are in the final treatment described in this chapter, a technique such as stepwise regression analysis must be used. Partial correlation coefficients are determined at every step in the stepwise procedure. The predictor variables are entered into the regression equation in order of the partial correlation coefficients with the effect of the common or shared variance removed.

The greatest degree of overlap is between MR and SR. Forty-

---

[52] Charles Spearman, <u>op. cit</u>., p. 5.

seven per cent of the aptitude measured is common to both tests. MR and SR both deal with diagrammatic items and both require a mental manipulation of the diagrammatic materials. It is not surprising, therefore, that they have a large amount of variance in common. AR and SR share twenty-seven per cent of their variance and again both tests deal with diagrammatic forms.

VR has substantial amounts in common with all of the other tests. It shares nineteen per cent of the variance with NA, twelve per cent with AR, twenty-three per cent with MR, and thirteen per cent with SR. If a test has a significant verbal content such common variance would be expected but the AR, MR, and SR tests all have minimal verbal content. This appreciable amount of common variance can perhaps be attributed to the "g" factor mentioned above. NA, like VR, shares common variance with the other tests, although to a lesser degree.

As mentioned, the common variance between the DAT means that they will not be independent predictors in the stepwise regression analysis. Further discussion of ways of compensating for this lack of independence is reserved for the last part of this chapter, in which the stepwise regression equations will be presented.

II. ANALYSIS OF THE RESULTS OF THE PHYSICS TESTS

Table III presents a summary of the results of the five physics tests. Three characteristics of the physics tests are reported; the reliability coefficient for each test, the level of difficulty for each test which is the fraction of students who responded correctly to the

TABLE III

LEVEL OF DIFFICULTY, BISERIAL CORRELATION AND RELIABILITY COEFFICIENTS FOR FIVE PHYSICS TESTS

Test	Mean Level of Difficulty	Mean Biserial Correlation	K - R 20 Reliability Coefficient	Mean	SD
PhV	.58	.54	.62(.76)	8.73	2.8
PhN	.54	.60	.59(.75)	6.43	2.5
PhG	.51	.55	.50(.75)	5.11	2.1
PhDV	.53	.56	.68(.81)	8.86	3.2
PhPD	.62	.58	.45(.71)	6.19	1.9

items in the test, and the biserial correlation which is the correlation between each test item and the total score for the test. The means, variances and standard deviations for each test are reported.

The tests are PhV--in which both stem and distractors are verbal; PhN--in which the stem is verbal and the distractors numerical requiring a numerical calculation; PhG--in which the stem is verbal and the distractors are graphs requiring numerical calculations and reasoning; PhDV--in which the stem of the item is verbal and contains a diagram and the distractors are verbal; PhPD--in which the stem is a diagram and the distractors are all diagrams and the verbal content is very limited and simple. More precisely, the biserial correlation is a measure of the relationship between the continuous variable (the total scores for all of the items of a particular physics test) and the dichotomous or two-categoried (right or wrong) variable.[53] The reliability coefficients for each of the five physics tests were determined by the use of the Kuder-Richardson Formula 20:

$$r = \frac{n}{n-1}\left[\frac{S^2 - \sum pq}{S^2}\right]$$

where r = reliability coefficient
n = the number of items in the test
S = the standard deviation of the distribution of scores for the physics test
p = the proportion of students who responded correctly to a given item
q = 1 - p [54]

---

[53]George A. Ferguson, op. cit., p. 199.

[54]Robert W. B. Jackson and George A. Ferguson, loc. cit.

The means and standard deviations for the five physics tests are also reported in Table III. For PhV the mean is 8.7 for a test of 15 items with a standard deviation of 2.8. For PhN the mean is 6.4 for a test of 12 items with a standard deviation of 2.5. For PhG the mean is 5.1 for a test of 10 items with a standard deviation of 2.1. For PhDV the mean is 8.9 for a test of 17 items with a standard deviation of 3.2. For PhPD the mean is 6.2 for a test of 10 items with a standard deviation of 1.8.

The mean scores for PhPD and PhV are larger in relation to the number of items than for any of the other tests. The mean level of item difficulty is also larger for PhPD and PhV than for the others (.62 and .58 respectively), and apparently these items are easier for the Grade XI students in the sample. The verbal items may be easier because they are a familiar type of item and because they require a single step--that of recall of information. The PhPD items are not familiar items, but they too require only one step in responding to them--that of a mental manipulation of the elements presented.

The mean biserial correlations of the items with the subtest total for each of the five physics tests are similar with PhN having the highest average biserial correlation (r_{bi} = .60). PhPD has very nearly the same average biserial correlation for individual items with its subtest total (r_{bi} = .58).

The PhPD test was composed of the easiest items, having a mean level of difficulty of .62, but it also had the second highest mean biserial correlation (r_{bi} = .58). PhV, on the other hand, had the

second highest mean level of difficulty (.58), but the lowest mean biserial correlation (r_{bi} = .54). The PhPD items were evidently somewhat more consistent in their correlation with the whole subtest than were the PhV items. Of the other three subtests, PhN had the largest mean biserial correlation (r_{bi} = .60). Even PhG, which had the lowest level of difficulty at .51, had a mean biserial correlation of .55.

The complete item analysis results which are presented in the Appendix (Table XIII), give an item by item analysis for each of the subtests. The distribution of levels of difficulty and biserial correlation is given in Table III. Items for the five physics subtests are fairly well distributed between the more difficult and the easier types of items. The biserial correlations between the items and the subtest totals are shown to be distributed over a range from .23 to .83 with forty-four of the sixty-four items having biserial correlations greater than .50.

An examination of the extremes of the distribution of levels of difficulty reveals that the eight most difficult items (levels of difficulty less than .30) had a mean biserial correlation of .49, and the seven least difficult items (levels of difficulty greater than .80) had a mean biserial correlation of .45. Even at the extremes, apparently the items performed reasonably well in discriminating among the students' achievement in physics.

The frequency distributions of the scores for the five physics subtests are presented in Figures 3, 4, 5, 6, and 7. Although a consideration of the means and standard deviations reported above gives essentially the same information, the presentation in graphic form

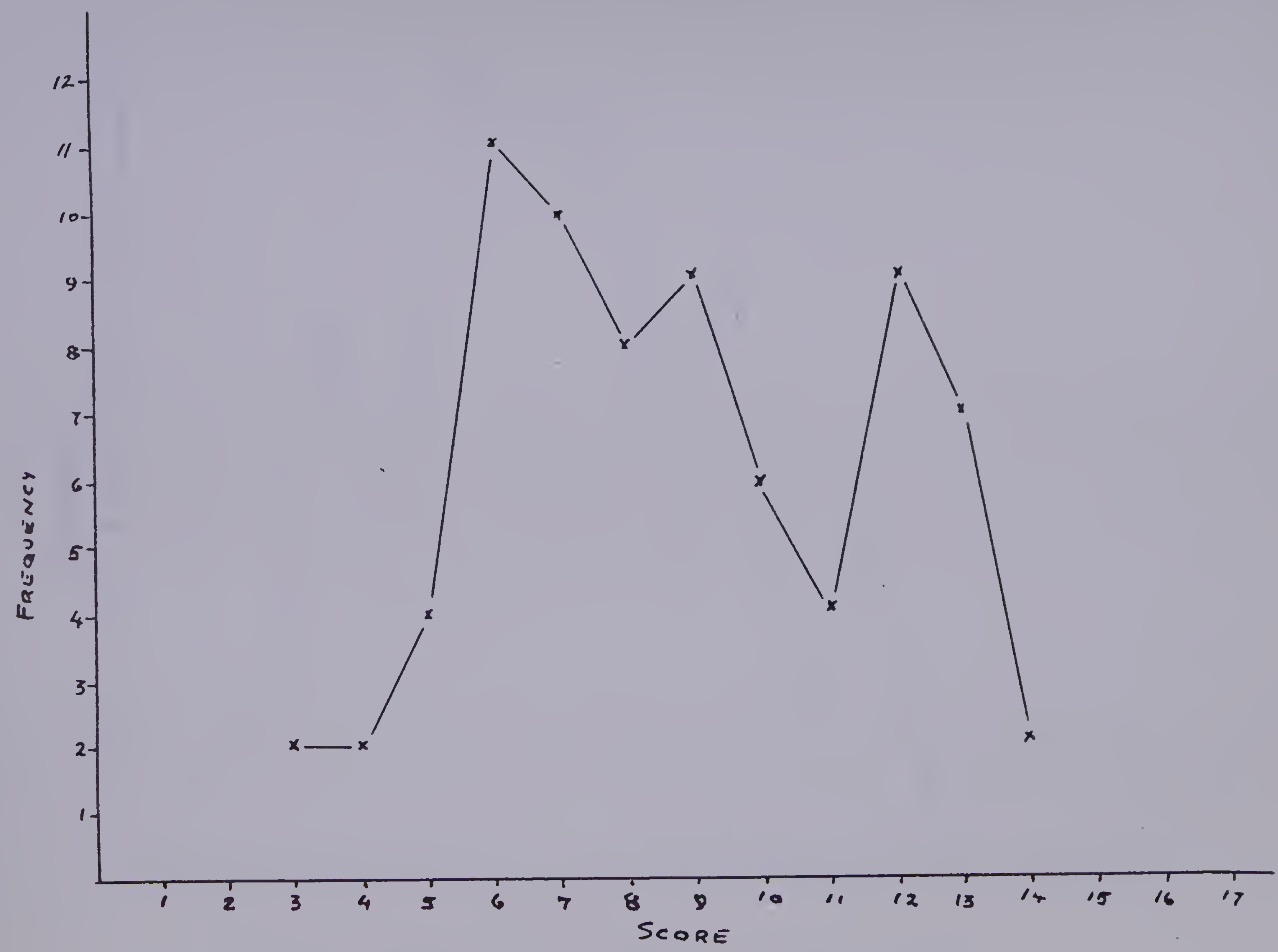


Figure 3. Frequency distribution of scores for PhV.

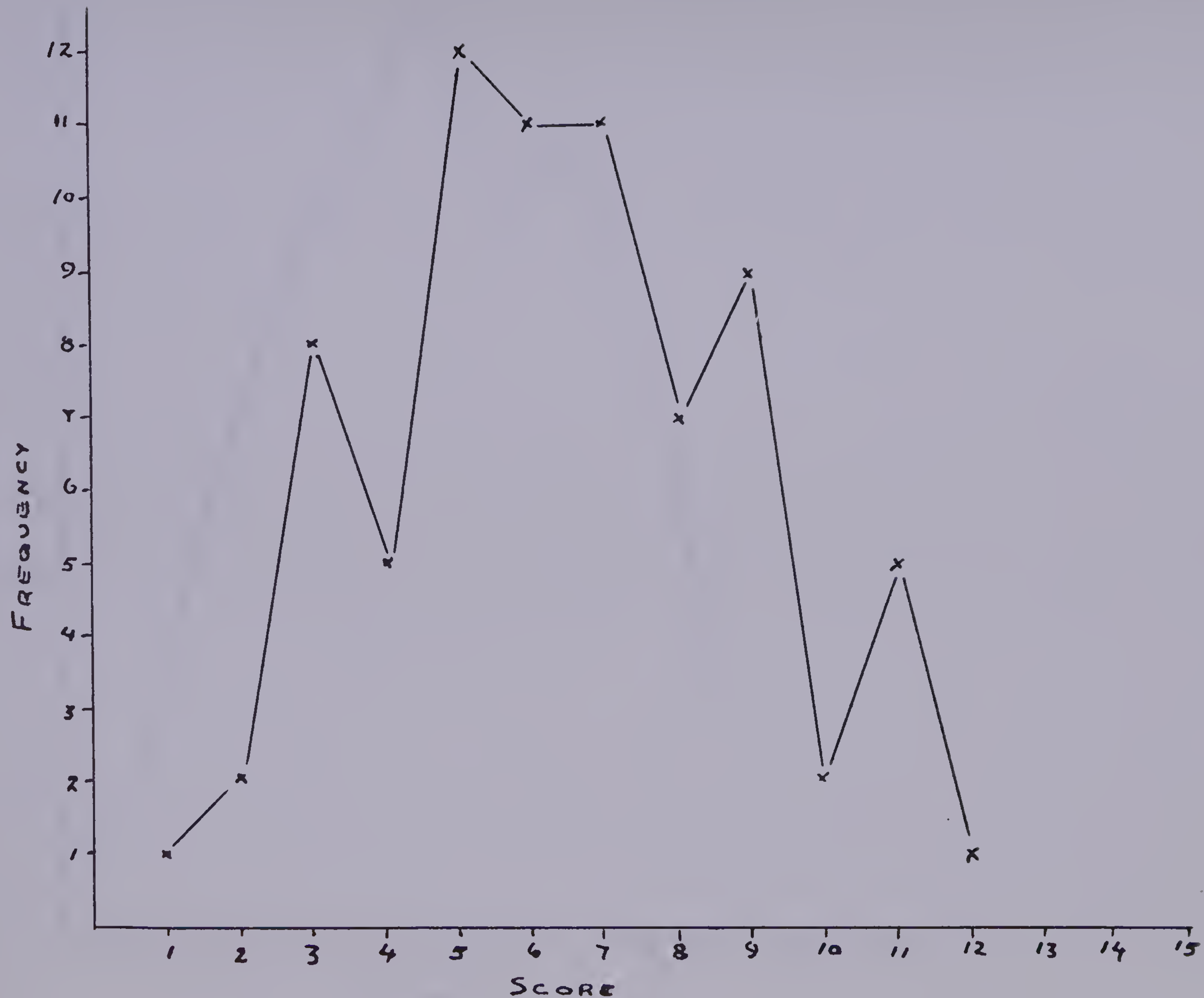


Figure 4. Frequency distribution of scores for PhN.

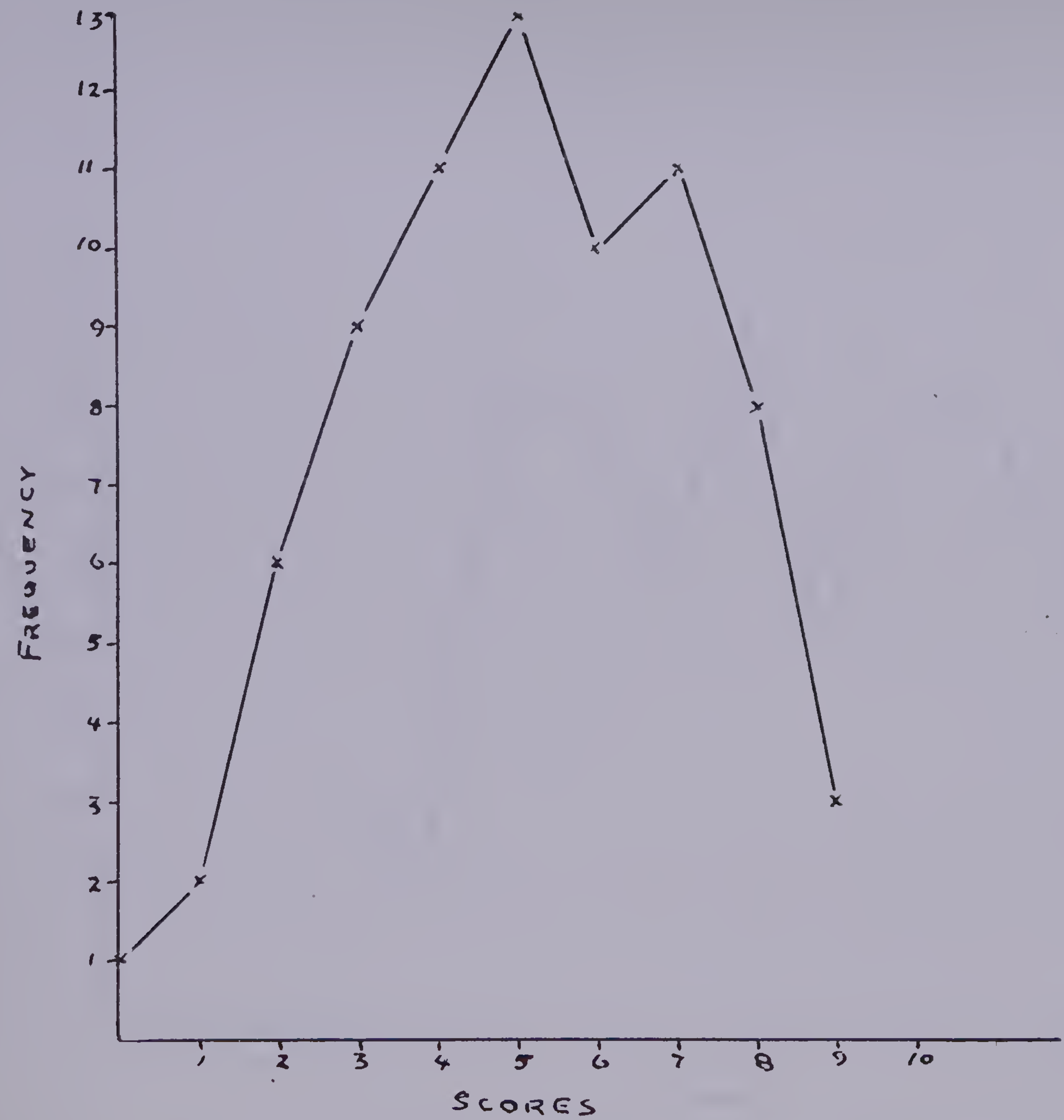


Figure 5. Frequency distribution of scores for PhG.

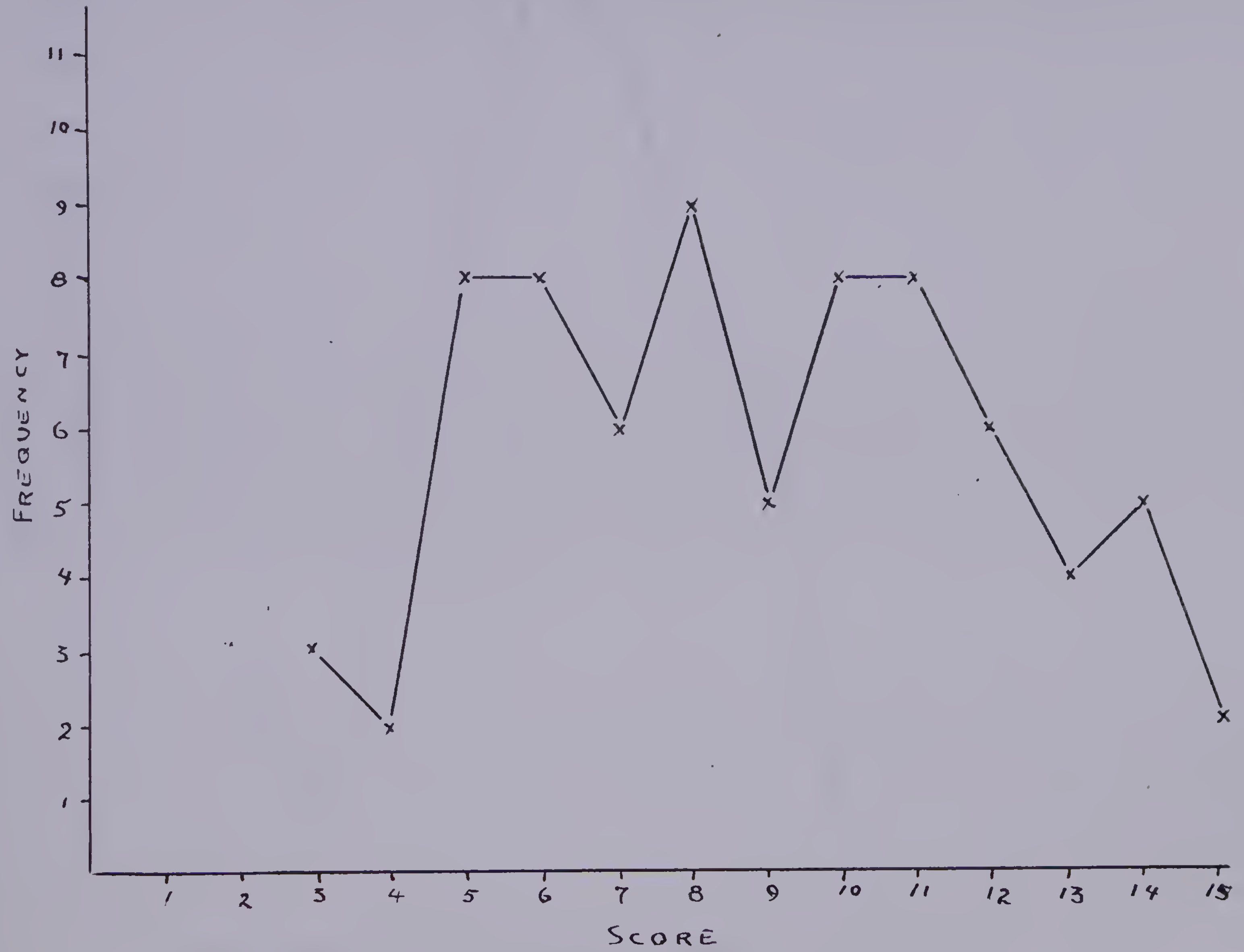


Figure 6. Frequency distribution of scores for PhDV.

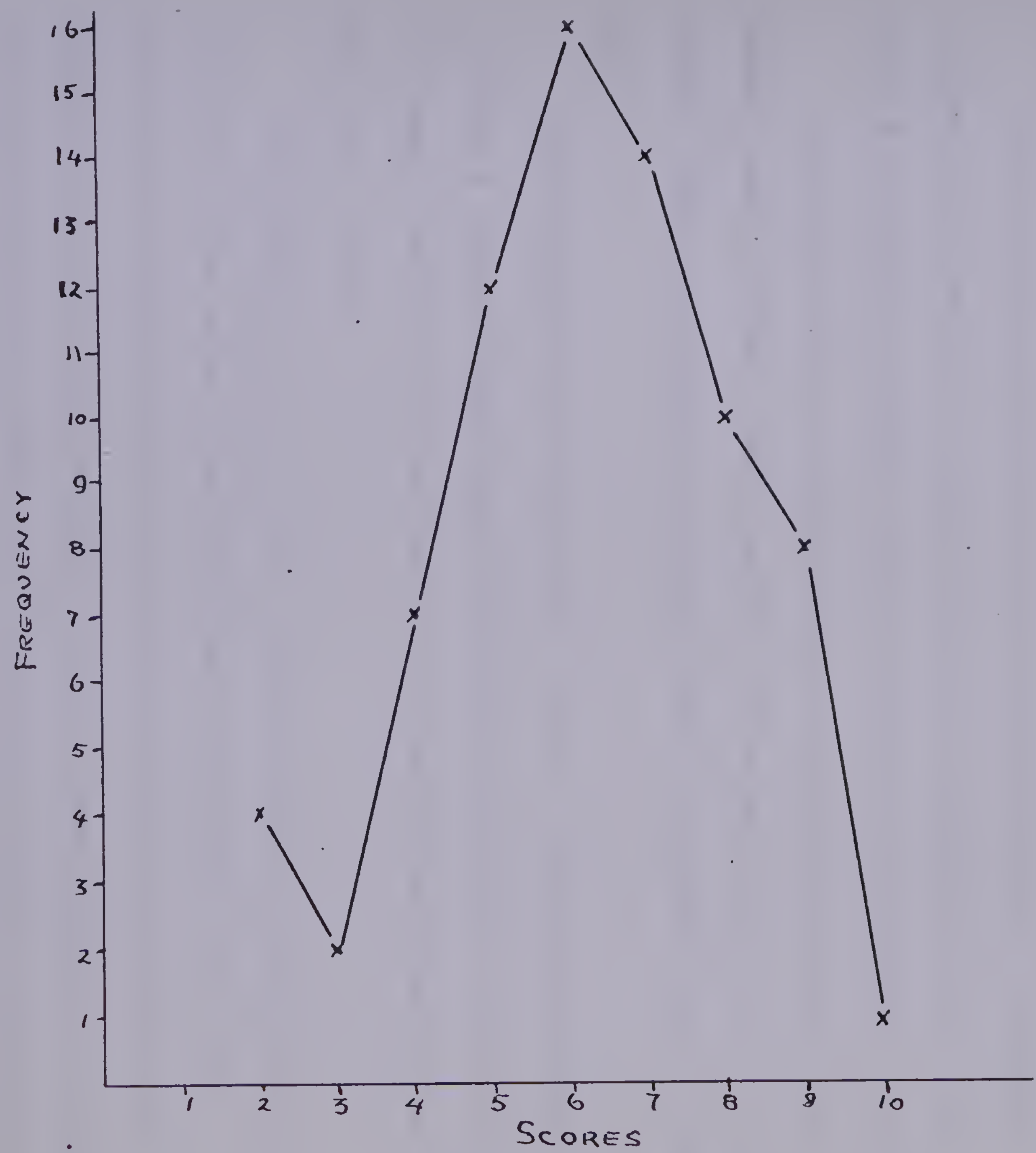


Figure 7. Frequency distribution of scores for PhPD.

serves to emphasize the differences among the five distributions.

In Figure 3 a pronounced bimodal distribution is observed for the PhV scores. However, the peaks of the two parts of the curve are reasonably close to each other since the SD of 2.8 for the test is fairly low.

The distribution of PhN scores as seen in Figure 4, shows an approximately bell-shaped distribution. The scatter of the scores is not abnormally large, the SD being 2.5.

The distribution of PhG scores as seen in Figure 5 is similar to that of the PhN test, although the curve is somewhat broader. The SD of 2.1 would also indicate that the scatter is not abnormally large.

The distribution of PhDV scores, as presented in Figure 6, is rather broad and flat. The SD is 3.2, the largest value for any of the tests. There are possibly not as many items of an easier nature in this test.

The distribution of PhPD scores, which is represented in Figure 7, shows a very pronounced peak with scores varying only a little from the mean. The SD is 1.8, smallest of the standard deviations for any of the physics tests.

The Figures 3, 4, 5, 6, and 7 indicate that the distributions of the physics test scores generally follow a bell-shaped curve. The scatter of the scores about the mean indicates that a reasonably good balance between difficult and easier items has been achieved. Combined with the information of Table III, and the more detailed item analysis presented in the Appendix Table XIII, the information presented in the frequency distribution graphs appears to substantiate the effectiveness

of the physics tests in discriminating among the students.

Reliabilities, as computed by the Kuder-Richardson Formula 20, were somewhat low. The values, as shown in Table III, were: for PhV, r = .62; for PhN, r = .59; for PhG, r = .50; for PhDV, r = .68, and for PhPD, r = .45. Three possible explanations of the comparatively low reliabilities may be suggested. One explanation has to do with the lengths of the five subtests. Long tests are more internally consistent, other things being equal, than are short tests.

Cronbach notes that with the addition of every question, the sample of performance becomes more adequate in terms of the inclusion of tests of all the kinds of performance which is being evaluated.[55]

The five physics tests have rather small numbers of items (PhV = 15, PhN = 12, PhG = 10, PhDV = 17, and PhPD = 10). If the three longer tests were doubled in size to PhV = 30, PhN = 24, PhDV = 34, and the two shorter tests were tripled in size to give PhG = 30 and PhPD = 30, the effect would be to give the five subtests roughly the same size. If, also, the source of the low reliability coefficients is due to test length alone, the Spearman-Brown prophecy formula

$$R = \frac{nr}{1 + (n - 1)r}$$

where R = the reliability coefficient for the extended test
r = the reliability coefficient for the test
n = the number of times the test length is extended[56]

---

[55]Lee J. Cronbach, Essentials of Psychological Testing (New York: Harper and Row, Publishers), 1970, p. 167.

[56]J. P. Guilford, Psychometric Methods (New York: McGraw-Hill Book Co. Inc.), 1954, p. 354.

may be used to predict the reliability coefficients for more extended tests.

The reliability coefficients for the five physics tests would then be:

PhV (15 items extended to 30) r = .75

PhN (12 items extended to 24) r = .75

PhG (10 items extended to 30) r = .73

PhDV (17 items extended to 34) r = .81

PhPD (10 items extended to 30) r = .71

Reliability coefficients in this range are much more acceptable for use as criteria of physics achievement in the multiple regression equations in the last part of this chapter.

A second possible source of the observed low reliabilities is guessing. Guessing may contribute to low reliabilities, where the number of items in the test is small as is the case in this study. Larger numbers of items might ensure that correct and incorrect guesses would balance one another. The raw scores for the physics tests are not corrected for guessing. This factor may have contributed to the large estimated error variance.

The third possibility is that the tests, particularly PhG and PhPD, are not composed completely of the same kind of item. If tests are not composed of the same kind of item they cannot, by definition, be internally consistent. If the students who respond correctly to other items do not respond correctly to a given item, then it is not consistent with the other items and the reliability of the test will be decreased. Awareness of this possibility must colour all conclusions

drawn from subsequent analyses of the data.

In all probability a combination of all three factors noted above has contributed to the low reliabilities of the physics subtests.

As Guilford comments regarding the assumption that intercorrelations of items are equal to the reliabilities of items:

> This is, of course, one of the conditions for a test measuring a single common factor. But this condition also probably applies when each item measures to about the same extent a weighted combination of more than one common factor. . . . If the common factors are distributed unevenly among the items, . . . the r_{tt} as estimated from the K - R formula will be reduced in size.[57]

For large numbers of items, this weighted combination approximates an average value, but for a smaller number of items, irregularities in single items can cause significant divergencies from the weighted combination referred to by Guilford. The PhPD items do vary in that one may have a numerical calculation, another requires the recalling of a physics experiment, and others may require a spatial visualization of the situation depicted in the diagram.

The intercorrelations among the physics tests are presented in Table IV. The fraction of the variance (expressed in per cent) shared by the tests is reported in parentheses following the correlation coefficients.

With the exception of the scores on PhPD, the intercorrelations are rather large. Since all five physics tests measure physics

---

[57] J. P. Guilford, op. cit., p. 383.

TABLE IV

THE MEANS, STANDARD DEVIATIONS, AND INTERCORRELATIONS
FOR THE PHYSICS TESTS
(N = 74)

Test	PhV	PhN	PhG	PhDV	PhPD
PhV					
PhN	.65*(42)				
PhG	.65*(42)	.56*(31)			
PhDV	.66*(43)	.67*(45)	.58*(34)		
PhPD	.43*(19)	.52*(27)	.28 (8)	.41*(17)	
Mean	8.7	6.4	5.1	8.9	6.2
Standard Deviation	2.9	2.5	2.1	3.2	1.9

Per cent of common variance - in parentheses following correlation coefficients

*Significant (p/0.05)

achievement, this was to be expected. The largest amount of common variance is shared by PhDV and PhN (45 per cent). In general, PhV, PhN, and PhDV account for the greatest amount of the shared variance.

Only PhPD appears to be markedly different in terms of the amount of variance it shares with the other physics subtests. Even PhG shares variance in amounts varying from 31 per cent to 42 per cent with PhV, PhN, and PhDV. The amount of variance shared by PhPD and the other four tests ranges from 8 per cent to 27 per cent, whereas the shared variance among the other four tests ranges from 31 per cent to 45 per cent.

This section has presented evidence about the five physics tests. Data reported on the tests includes the level of difficulty and the biserial correlation for each item as well as the means, standard deviations, frequency distributions, and reliability coefficients for each of the five physics subtests.

The item analysis data, the means, the standard deviations, and the frequency distributions would appear to indicate that the tests were composed of items of varying degrees of difficulty. The items appear to discriminate adequately among the students in the sample.

Approval of the reliabilities of the tests must be much more guarded. Reliability coefficients, as determined by the K - R Formula 20 for the five physics subtests, were low. If, however, the low reliabilities were due only to the relatively small number of items in the tests, a correction using the Spearman-Brown prophecy formula would yield reliability coefficients in excess of .70 for the extended tests. If the tests are reliable to the extent indicated by the reliability coefficients for the extended tests, they might be adequate for use as

criterion variables in the regression analysis.

The rather large intercorrelations which were observed among the physics subtests indicate that they, like the DAT, all have a large amount of variance which is common to more than one test. The regression equations which result from the stepwise regression analysis must be considered separately because the equations so determined overlap. However, since the purpose of this study was to measure the effect of changing the amount and kind of verbal and diagrammatic content in the physics test items, the five separate regression analyses were carried out.

III. INTERCORRELATIONS AMONG THE PHYSICS TESTS AND THE D A T

The intercorrelations among the physics tests and the DAT subtests are presented in Table V. The hypotheses A - E were tested using the intercorrelations presented in Table V.

A. Physics-verbal (PhV)

Hypotheses A(1), A(2), A(3), A(4), and A(5), which state that no significant (at p/0.05) zero-order correlations will be observed between PhV and VR, NA, AR, MR, and SR were tested by computing intercorrelations among the physics tests and DAT subtests. All of the hypotheses A(1) to A(5) must be rejected because zero-order correlations greater than the critical value .23[58] were observed.

The intercorrelations between PhV and the DAT subtests were all

[58] G. A. Ferguson, Statistical Analysis in Psychology and Education (New York: McGraw-Hill Book Co. Inc.), 1966, p. 413.

TABLE V

INTERCORRELATIONS AMONG THE D A T SUBTESTS AND THE PHYSICS TESTS

	VR	NA	AR	MR	SR
PhV	.57*(32)	.50*(25)	.30*(9)	.41*(17)	.33*(11)
PhN	.52*(27)	.47*(22)	.26*(7)	.45*(20)	.45*(20)
PhG	.56*(31.5)	.40*(16)	.10 (1)	.44*(19)	.30*(9)
PhDV	.55*(30)	.40*(16)	.34*(12)	.55*(30)	.51*(26)
PhPD	.37*(14)	.29*(8)	.17 (2.9)	.43*(19)	.42*(18)

Per cent of the variance shared by the tests given in parentheses following the correlation coefficient.

*Significant at $p < .05$

observed to be significant at p/0.05. The greatest amount of the variance of the PhV is shared with VR--32 per cent. The PhV test also shares 25 per cent of the variance with NA. This observation, although perhaps not expected on logical grounds, was not unexpected on the grounds of experience with the DAT. The DAT authors state that VR and NA are good indicators of academic success,[59] and verbal items are characteristically academic types of items. The authors of the DAT conjecture that VR and NA measure a general learning ability.[60]

B. Physics-numerical (PhN)

Hypotheses B(1), B(2), B(3), B(4), and B(5), which state that no significant (at p/0.05) zero-order correlations will be observed between PhN and VR, NA, AR, MR, and SR were tested by computing intercorrelations between PhN and each of the DAT subtests. All of the hypotheses B(1) to B(5) must be rejected because zero-order correlations greater than the critical value .23 were observed.

In general, the intercorrelations which were observed between PhV and the DAT subtests and those between PhN and the DAT subtests were very much alike. Although AR was found to be significantly correlated with PhN, it was only marginally correlated. PhN shared only 7 per cent of the variance with AR. The PhN test shared 27 per cent of the variance with VR. The amount of variance shared by PhN and MR, as well as that shared by PhN and SR, was greater, being 20 per cent in

---

[59] George K. Bennett, Harold G. Seashore, Alexander G. Wesman, *op. cit.*, p. 3-5.

[60] *Ibid.*, p. 5-3.

both cases, than that shared by PhV and the same DAT subtests. The tasks of visualizing the apparatus or the objects described in the question stem may have been more important in the numerical items than they were in the verbal items.

C. Physics-graphs (PhG)

Hypotheses C(1), C(2), C(3), C(4), and C(5), which state that no significant (at p/0.05) zero-order correlations will be observed between PhG and VR, NA, AR, MR, and SR, were tested by computing inter-correlations between PhG and each of the DAT subtests. Hypotheses C(1), C(2), C(4), and C(5) must be rejected because zero-order correlations greater than the critical value .23 were observed. Hypothesis C(3) cannot be rejected because AR shares only 1 per cent of the variance with PhG.

Although the verbal reasoning subtest of the DAT was observed to be the subtest most highly correlated with PhG, sharing 31.5 per cent of the variance, MR was found to be the next most highly correlated variable replacing NA. The MR subtest shared 19 per cent of the variance with PhG. It appears possible that after the verbal interpretation of the item stem, the next important task which confronted the student was that of imagining or visualizing the apparatus before the numerical relationships between the variable were determined and interpreted graphically.

D. Physics-diagrammatic verbal (PhDV)

Hypotheses D(1), D(2), D(3), D(4), and D(5), which state that no significant (at p/0.05) zero-order correlations will be observed between

PhDV and VR, NA, AR, MR, and SR, were tested by computing intercorrelations between PhDV and each of the DAT subtests. All of the hypotheses must be rejected because zero-order correlations greater than the critical value .23 were observed.

Both VR and MR were observed to share 30 per cent of the variance with the PhDV test. SR shared 26 per cent of the variance with PhDV. Because of the dependence of these items on understanding the diagrams which accompanied them, both MR and SR were significantly correlated with PhDV. Where there was verbal content, however, the VR subtest remained perhaps the best predictor for PhDV test scores.

E. Physics-pure diagrammatic (PhPD)

Hypotheses E(1), E(2), E(3), E(4), and E(5), which state that no significant (at $\underline{p}/0.05$) zero-order correlations will be observed between PhPD and VR, NA, AR, MR, and SR, were tested by computing intercorrelations between PhPD and each of the DAT subtests. Hypotheses E(1), E(2), E(4) and E(5) must be rejected because zero-order correlations greater than the critical value of .23 were observed. Hypothesis E(3) cannot be rejected because the intercorrelation of PhPD and AR was smaller than .23. These two variables share only 3 per cent of the variance.

The levels of intercorrelation between PhPD and the five DAT subtests were lower than for the other four physics tests. The fraction of variance shared by PhPD and MR, SR, VR, and NA ranged between 19 per cent with MR to 8 per cent with NA. The PhPD test was the only one of the five physics tests for which VR was not the best predictor.

The first three parts of this chapter have included examinations

of the DAT subtests and the physics tests. The intercorrelations between the DAT and the physics tests were examined and a more detailed and searching analysis of the relationship will be carried out by means of the technique known as stepwise regression analysis.

IV. THE STEPWISE REGRESSION ANALYSIS

The final section of this chapter presents a more detailed analysis of the relationship between the physics tests and the DAT subtest predictors. The analysis was carried out by the "stepwise regression" procedure described in Chapter III. As indicated in the previous chapter, this procedure consists of successive approximations to a multiple linear regression equation of the form:

$$Y = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3 + \ldots + e$$

in which Y = the distribution of scores on the physics tests,
$X_1, X_2, X_3, \ldots$ = the distributions of the scores on the DAT subtests
b_0 = a constant
$b_1, b_2, b_3, \ldots$ = the coefficients which produce the 'best fit.'
e = the error

The analysis was carried out by means of the stepwise regression program (MULR 06) made available to the investigator by the Research and Planning Branch of the Manitoba Department of Youth and Education, who obtained it from the Faculty of Education at the University of Alberta.

The final best regression equation was arrived at through a series of steps in which successive predictor variables were entered into the regression equation in order of the magnitudes of their

partial correlations with the criterion variable (the physics test scores).

At every step of the stepwise procedure, an F ratio, distributed with n and N - n - 1 degrees of freedom, was generated (where n = the number of predictor variables in the regression equation, and N = the number of students).

The decision to retain a variable in the regression equation was based on the probability of less than 0.05 of observing a larger F ratio.

A. Physics-verbal (PhV)

On the basis of the data presented in Table VI, hypothesis F, namely that no significant improvement to the improvement of predictions of PhV scores will be made by adding to the Verbal Reasoning predictor measures of Numerical Ability, Mechanical Reasoning, Space Relations, and Abstract Reasoning, was rejected.

From Table VI it may be seen that both VR and NA make significant (at $p \angle 0.05$) contributions to the prediction of scores on the PhV test. When both predictor variables have been entered into the regression equation, 39.9 per cent of the variance of the PhV scores was accounted for. With both predictors in the equation, a total F value of 23.6 was observed. The F value for the last variable entering the equation was 9.2, and the probability of observing a larger F value was .001.

This combination of variables might have been predicted a priori from the evidence presented by the authors of the DAT, that VR + NA generally constitute the best predictor of academic success. As the sample of students has been demonstrated to be academically able, this

TABLE VI

STEPWISE PREDICTION OF PHYSICS-VERBAL SCORES (PhV)
FROM THE D A T SUBTESTS

	Predictor Variable Entering	F Value For Last Variable Entering	Total F Value	Probability Level For Last Variable	Percent Of Variance Accounted For By Significant Variables
Step 1	VR	34.1	34.1	$\underline{/}0.001$	32.1
Step 2	NA	9.2	23.6	$\underline{/}0.001$	39.9

Best Regression Equation:

at $p\underline{/}0.05$ $\quad Y_{PhV} = .22\ X_{VR} + .17\ X_{NA} - 5.25$

result is not surprising.

The item type in PhV was one familiar to these students. The items required the recall of facts and terms, a task at which academically successful students have been generally adept throughout their school careers.

B. Physics-numerical (PhN)

On the basis of the data presented in Table VII, hypothesis G, namely that no significant improvement will be made to the prediction of scores on the PhN test by adding to the Verbal Reasoning predictor measures of Space Relations, Numerical Ability, Mechanical Reasoning, and Abstract Reasoning was rejected.

From Table VII it may be seen that VR, SR, and NA all make significant (at $p \angle 0.05$) contributions to the prediction of scores on the PhN test. When all three predictor variables have been entered into the regression equation, 39.2 per cent of the variance of the PhN scores was accounted for. With the three predictor variables in the regression equation, a total F value of 15.0 was observed. The F value for the last predictor (NA) entering the equation was 5.4. The probability of observing a larger F value was 0.02.

Although it might have been expected that the NA predictor would make an earlier entry into the regression equation and thus have a greater effect on it, this was not the case. Apparently the first and most important task confronting the student in responding to numerical items is that of understanding the verbal content. A further explanation may be found in the conjecture that the VR test measures a general

TABLE VII

STEPWISE PREDICTION OF PHYSICS-NUMERICAL SCORES (PhN) FROM THE D A T SUBTESTS

	Predictor Variable Entering	F Value For Last Variable Entering	Total F Value	Probability Level For Last Variable	Percent of Variance Accounted For By Significant Variables
Step 1	VR	26.3	26.3	<0.001	26.8
Step 2	SR	8.4	18.7	<0.001	34.5
Step 3	NA	5.4	15.0	0.02	39.2

Best Regression Equation:

at $p \angle 0.05$ $\quad Y_{PhN} = 0.14\ X_{VR} + 0.12\ X_{SR} + 0.07\ X_{NA} - 5.55$

learning ability.

Whether VR measures "g" or not, it is not unreasonable that VR and SR should be significant predictors of PhN. The stems of the item are somewhat complex verbally and the items also frequently require that a student be able to visualize the position of a piece of apparatus. It is therefore not unexpected that verbal reasoning and spatial tests should be significant predictors of success. The computational process by which the answer is obtained is the least influential of the significant predictors.

C. Physics-graphs (PhG)

On the basis of the data presented in Table VIII, hypothesis H, namely that no significant improvement will be made to prediction of scores on the PhG test by adding to the Verbal Reasoning predictor measures of Mechanical Reasoning, Numerical Ability, Space Relations, and Abstract Reasoning, was rejected.

From Table VIII it may be seen that VR, MR, and SR all make significant (p/0.05) positive contributions to the prediction of PhG scores and, further, that NA makes a significant negative contribution. With all four predictor variables in the regression equation, the total amount of variance accounted for was 41.0 per cent. The total F value observed when all four predictors were entered in the regression equation was 12.0. The F value for the last predictor entering the equation (AR) was 4.1. The probability of observing a larger F value was 0.047. For the third step, the total F value was 14.0, the F value for the last variable entering the equation was 2.6, and the probability

TABLE VIII

STEPWISE PREDICTION OF PHYSICS-GRAPH SCORES (PhG)
FROM THE D A T SUBTESTS

	Predictor Variable Entering	F Value For Last Variable Entering	Total F Value	Probability Level For Last Variable	Percent Of Variance Accounted For By Significant Variables
Step 1	VR	33.1	33.1	0.00	31.5
Step 2	MR	4.1	19.3	0.046	35.3
Step 3	NA	2.6	14.0	0.11	37.6
Step 4	AR	4.1	12.0	0.047	41.0

Best Regression Equation:

at $p \leq 0.05$ $\quad Y_{PhG} = .16\,X_{VR} + .09\,X_{MR} - .11\,X_{NA} + .08\,X_{AR} - 3.75$

of observing an F value larger than 2.6 was 0.11. The phenomenon observed in this analysis has been called the effect of a suppressor variable. NA makes a small positive, but not a significant, contribution to the prediction of the PhG test scores. The observed effect of adding AR as a predictor is to produce a regression equation in which the coefficient for the NA variable becomes negative. NA is apparently a suppressor variable in the regression equation.

Under the best conditions for the prediction of test scores by multiple regression techniques, there should be high correlation between the criterion and each of the predictor variables, but low intercorrelation among the predictors. The predictors ought to be independent in other words. Such is not the case in this instance because the DAT subtests have rather large intercorrelations.

The conditions observed in this analysis conform with Lord and Novids' description of a suppressor variable:

> Variables that are useful in a regression equation because they have low or zero correlation with the criterion, but high correlation with other predictors, are called <u>suppressor variables</u>. In effect, they represent some aspect of the predictor variables that is not related to the criterion but that function to 'suppress' or subtract out this invalid component and thus to make the original predictor more valid.[61]

In the prediction equation for PhG, the NA predictor, when added

---

[61] Frederic M. Lord and Melvin R. Novick, <u>Statistical Theories of Mental Test Scores</u> (Reading, Mass.: Addison-Wesley Publishing Co.), 1968, pp. 27-28.

to the equation by itself, makes a small positive, but not a significant, ($p \underline{/} 0.05$) contribution to the equation. When AR is added, however, the NA predictor makes a significant but negative contribution. In explaining this phenomenon it must be remembered that the PhG items were often quite difficult requiring the derivation of some data from which a graph may be constructed. Whereas the process of actually computing the data may have some minimum effect on the ability to respond to the item, the real difficulty lies in reasoning what calculations to perform--depending upon an abstract reasoning ability. Thus the addition of the AR predictor may make the process of computation a negative predictor of success in PhG items.

The other significant predictors seem reasonable when the nature of the PhG items is recalled. The items are highly verbal and thus the first and most significant predictor is VR. MR produces a significant positive contribution because in many items a degree of understanding of a mechanism is required. For example, the relative positions of the object, lens, and images in items 33, 34, and 35 require an ability to comprehend the working of this mechanism as the object changes its position with respect to the lens.

D. Physics-diagrammatic verbal (PhDV)

On the basis of the data presented in Table IX, hypothesis I, namely that no significant improvement will be made to the prediction of scores on the PhDV test by adding to the Verbal Reasoning predictor measures of Space Relations, Mechanical Reasoning, Numerical Ability, and Abstract Reasoning, was rejected.

TABLE IX

STEPWISE PREDICTION OF PHYSICS-DIAGRAMMATIC VERBAL SCORES (PhDV) FROM THE D A T SUBTESTS

	Predictor Variable Entering	F Value For Last Variable Entering	Total F Value	Probability Level For Last Variable	Percent Of Variance Accounted For By Significant Variables
Step 1	VR	31.1	31.1	<0.001	30.2
Step 2	SR	13.0	24.7	<0.001	41.0

Best Regression Equation:

at $p \leq 0.05$ $\quad Y_{PhDV} = .24\ X_{VR} + .12\ X_{SR} - 5.22$

It may be seen from Table IX that both VR and SR contribute significantly ($p < 0.05$) to the prediction of the PhDV test scores. The total F value for the combination of predictors was observed to be 24.7. The total amount of variance accounted for by the two predictor variables was 41.0 per cent. The F value for the last variable (SR) which entered the regression equation was 13.0. The probability of observing an F value larger than 13.0 was less than 0.001.

The first variable to enter the multiple regression equation was VR. The nature of the items was such that the verbal skills were most useful to the student in responding to the PhDV items. Many of the instructions were verbally quite complex. The required verbal reasoning and the general learning ability or "g" associated with VR probably both operated to make VR the best predictor. A mental manipulation of the objects or apparatus represented in the diagrams was also required in order to respond correctly to the items.

E. *Physics-pure diagrammatic* (PhPD)

On the basis of the data presented in Table X, hypothesis J, namely that no significant improvement will be made to the prediction of scores of the PhPD test by adding to the *Mechanical Reasoning* predictor measures of *Verbal Reasoning*, *Space Relations*, *Numerical Ability*, and *Abstract Reasoning*, was not rejected.

The MR subtest accounted for 18.8 per cent of the variance of the PhPD test scores. The F value for the MR predictor was observed to be 16.7. The probability of observing a larger F value was less than 0.001.

The only significant predictor of PhPD scores was observed to be

TABLE X

STEPWISE PREDICTION OF PHYSICS-PURE DIAGRAMMATIC SCORES (PhPD) FROM THE D A T SUBTESTS

	Predictor Variable Entering	F Value For Last Variable Entering	Probability Level For Last Variable	Percent Of Variance Accounted For By Significant Variables
Step 1	MR	16.7	<0.001	18.8

Best Regression Equation:

at $p \leq 0.05$ $\quad Y_{PhPD} = .12\ X_{MR} - 0.10$

MR. The appearance of MR as the only significant predictor from the DAT subtests of success in PhPD items suggests that the understanding of what is happening in the real situation which the diagrams represent may be unrelated to a verbal understanding of the events. If one of the objectives of evaluation in physics is that this kind of non-verbal understanding is useful and to be promoted, the achievement of non-verbal understanding can be evaluated by the use of pure diagrammatic items. It may be that some students can better understand test items which are constructed as pure diagrammatic types.

The results of the stepwise regression analysis are summarized in Table XI. The significant predictors from the DAT battery for each of the physics tests are marked by an asterisk.

Comparisons of the regression equations must be very guarded and qualified because of the high intercorrelations among the criterion variables (the physics tests). It is very likely that the five regression equations overlap in the sense that they all may describe some of the variance which is common to all of the tests. The appearance of VR as the first predictor in all but one of the regression equations might lead one to conclude that it does measure a general learning ability factor or "g" in addition to a specific verbal ability factor. It seems possible that in at least some of the regression equations, the VR predictor functions in both of these ways.

The relationship between verbal and diagrammatic content in physics test items is of particular interest in this study, and a consideration of the PhDV and PhPD variables may be of interest. Two

TABLE XI

SIGNIFICANT D A T PREDICTORS OF PHYSICS TEST SCORES

Test	VR	NA	AR	MR	SR
PhV	*	*			
PhN	*	*			*
PhG	*	*	*	*	
PhDV	*				*
PhPD				*	

predictors make significant contributions to the prediction of PhDV, namely VR and SR. Only one predictor variable, MR, makes a significant contribution to the prediction of PHPD test scores. If VR functions partly as a measure of "g," that part is not large enough to contribute significantly to the prediction of PhPD test scores. In the PhDV items where there is sometimes a rather complex verbal component to the item, the VR predictor is significant. Whether this verbal predictor is due solely to the complexity of the verbal component of the test items, or whether it is due to a combination of "g" and verbal reasoning, may be open to question. A verbal component is, however, measureable in the PhPD items and although the component is not a significant predictor, it would have been the second predictor to enter the regression equation.

It is perhaps unfortunate that both Mechanical Reasoning and Space Relations were used as predictors in the stepwise analysis. The MR and SR variables share so much variance ($r_{MR-SR} = .67$) that they can almost be used interchangeably in the regression equation. On the four occasions when one or the other are significant predictors in the regression equation, twice it was the MR predictor which entered the equation and twice it was SR. The combination of MR and SR does not appear in any of the regression equations.

The Numerical Ability predictor appears in the PhV, PhN, and PhG regression equations. The appearance of the VR and NA combination has already been noted and commented upon in relation to the best regression equation for the prediction of the PhV scores. Its function in that particular equation may be a part of a measure of general learning ability. In the PhN and PhG regression equations, the NA predictor may

be ambiguous in that part of the variance it explains may be due to "g" and part to the specific computational skills required in responding to the numerical and graphic items.

The only appearance of Abstract Reasoning in the regression equation for PhG appears to have been in connection with a suppressor variable. The interpretation of the role of AR in that equation is rather doubtful.

CHAPTER V

SUMMARY AND RECOMMENDATIONS

The final chapter summarizes the results of the investigation and seeks implications for teaching and testing in secondary school physics. Some considerations for course selection by students are suggested. Finally, a number of recommendations for further research are suggested.

I. SUMMARY

The purpose of the study was to determine the relationship between student achievement on physics test items in which the amount and kind of verbal and diagrammatic content varied systematically and certain intellectual aptitudes as they were measured by the DAT. The physics tests were of five different kinds. The first was a verbal test of physics in which both the stem of the items and the distractors were in verbal form. The second was a numerical test of physics for which the stem for each item contained both verbal and numerical information and the distractors were all numerical. The third was a graphic test in which the stem of the items were verbal and the distractors were graphs. The fourth was a diagrammatic-verbal test in which the stems were partly verbal and also contained a diagram. All the distractors for the diagrammatic-verbal items were verbal. The fifth test was pure diagrammatic in form and the stem and the distractors for this test consisted of diagrammatic representations. Any verbal instructions necessary were kept very simple and in the fewest possible words. The physics tests and the DAT sub-

tests VR, NA, AR, MR, and SR were administered to 74 students at Vincent Massey Collegiate in Winnipeg.

The results of the physics tests were item-analyzed. The reliability coefficient was computed for each of the five physics tests and the biserial correlation was determined for each item. The intercorelations among the DAT subtests, among the physics tests, and between each physics test and each of the DAT subtests were also determined. The relationship between physics achievement and the DAT was examined by the stepwise regression procedure, which was employed in order to determine the best set of DAT predictors for each of the five physics tests.

The results of the study showed that the subjects were above average in ability in all of the aptitudes tested. Their superiority was particularly evident in the VR, NA, and AR tests. This was not surprising because the students were in an academic course stream and had elected PSSC Physics.

The intercorrelations among the DAT were rather large, all of them being significant at $p \leq 0.05$. The DAT subtests are, therefore, not independent predictors of well-defined and distinct aptitudes, but each test has a substantial amount in common with each of the other tests.

The item analysis of the physics tests produced somewhat low reliability coefficients which cast a degree of doubt on the generalizations which have based on the results of this study. The low reliability coefficients may be a consequence of the limited numbers of items in each of the tests. The Spearman-Brown prophecy formula yielded reliabilities in excess of .70 for each test if expanded to approximately

30 items.

Biserial correlations between the items of the physics tests and the scores on each test were generally fairly high. The item difficulties were fairly well distributed between difficult and easier items. Because of the evidence of the biserial correlation and level of difficulty, it appears that the tests discriminate fairly consistently between the better students and the poorer ones.

The intercorrelations among the physics tests were generally high, except for those between PhPD and the other four tests. PhPD (the pure diagrammatic items) was found to share as little as eight per cent of the variance with PhG. The largest amount of shared variance for PhPD is 27 per cent with PhN. On the other hand, the smallest amount of variance shared by any other pair of physics tests is 31 per cent between PhN and PhG. The largest value, between PhDV and PhN, is 45 per cent.

The intercorrelations between the DAT and the physics tests revealed a wide variation in the degree of relationship. As little as one per cent of the variance was shared between PhG and AR. At the other extreme, PhV and VR had 32 per cent of their variance in common.

The stepwise regression analysis revealed some interesting relationships between the physics tests and the aptitude tests. The significant predictors in the multiple regression equations for each of the physics tests are presented in Table XI in the preceding chapter.

The significant predictors for PhV (verbal items), accounting for 40 per cent of the criterion variance, were VR and NA. The significant predictors for PhN (numerical items), accounting for 39 per cent of the variance of the criterion scores, were VR, SR, and NA. The significant

predictors for PhG (graphic items), accounting for 41 per cent of the variance of the criterion scores, were VR, MR, NA, and AR. In the "best regression equation" the NA predictor was found to have a negative coefficient. For the PhDV (diagrammatic-verbal items), the significant predictors, accounting for 41 per cent of the variance of the PhDV scores, were the VR and SR predictors. Finally, the only significant predictor for the PhPD (pure diagrammatic items) was MR, and it accounted for only 19 per cent of the variance of the criterion scores.

It is noted that the first predictor to enter the regression equation, in every case except for PhPD, was VR. Apparently where the verbal content of any test is appreciable, the task of interpreting this verbal content is the most significant one.

There appear to be three fairly distinct predictors which enter the multiple regression equations. The three are VR, NA, and the combination MR and SR. It may be justifiable to think of a predictor combining MR and SR because of their large intercorrelation (.67), and also because of psychological theories which place mechanical and spatial abilities in a closely related cluster. One such theory is that of Vernon to which reference has been made in Chapter II and which may be seen in Figure 1, page 11.[62] MR and SR would measure, presumably, a portion of the ability which Vernon refers to as the k:m factor.

The AR (Abstract Reasoning) predictor was found to enter only one of the regression equations and then in a relatively minor role.

---

[62]Phillip E. Vernon, The Structure of Human Abilities (London: Methuen), 196 , p. 72.

II. RECOMMENDATIONS

One of the important differences which exist between students is the difference in their pattern of aptitudes. If teachers are to truly recognize the differences between students, perhaps achievement should be tested with this in mind. It is perhaps not too far fetched to suggest that each student may require a different kind of test in order to evaluate the student's understanding adequately and fairly. The test would, ideally, allow a student to demonstrate understanding in the particular manner most suited to that student.

Teachers, in constructing tests and examinations, should be aware not only of the demands of the subject matter, but also of the variety of aptitudes which students have. Tests ought to be a representative sample of the subject matter and should also sample in a representative manner the aptitudes which students have. It is possible to frame items designed to measure the understanding of physics in such a complex verbal form as to eliminate the possibility of a correct response for some students. If the same concept can be tested by a diagrammatic item there is some evidence in this study that it ought to be so tested.

Some reasons for the foregoing statements may be inferred from two aspects of the study. First, it was demonstrated that the students in the sample differed measurably from the standardizing population in their aptitudes. Second, the physics tests were predicted best by different sets of DAT subtests as summarized above in Table XI. It may perhaps be inferred from the regression equations that, if the verbal content of a test item is appreciable, the task of interpreting this verbal content

is the first significant one confronting the student. The degree to which the verbal reasoning factor was dependent upon "g" is open to question. Nevertheless, in the PhPD tests, in which the verbal content was simple and minimal, the VR predictor was not found to be significant. If, therefore, a concept which depends on a visual or spatial understanding is to be tested, it should be tested with items which are as free from complex verbalism as possible.

It is obvious that some conceptions of physics are verbal, the knowledge of terminology, the statement of some fundamental laws are verbal, and knowledge of them can only be meaningfully tested in a verbal context. The argument here is that concepts which depend upon visual or spatial understanding may be better measured by pure diagrammatic items. Such items are difficult to construct but it may be worthwhile to spend the additional time and effort if these items identify a different aspect of understanding than is usually identified by conventional verbal items. It appears to be reasonable that pure diagrammatic items are simpler to use because a verbal description of some piece of apparatus or of some object may require words, whereas a diagram may describe it more economically and more accurately.

III. RECOMMENDATIONS FOR FURTHER RESEARCH

The following recommendations are made for further research in order to extend or refine the findings of this study.

1. Improve the tests by increasing the number of items and further refining the wording so that the test reliabilities may be increased. Extend the scope of the questions so that the levels of

understanding as defined by Bloom may be identified as a further variable in the study.

2. Replicate the study using as subjects students in a non-academic or general stream and using the content of physics as it appears in the general science curriculum for these students.

3. Replicate the study with a parallel forms test in which the same concepts are tested by verbal, numerical and diagrammatic types of items.

4. Conduct a study of the effect upon student achievement of the increasing verbal complexity of test items.

5. Extend the study to include items based on the interpretation of motion pictures. Since diagrams are intended to portray, in two dimensions, the reality of a three-dimensional situation, it may be that an extension from a two-dimensional but static representation of reality to a two-dimensional and dynamic representation will improve the ability of some students to respond correctly to the test items.

6. Conduct research aimed at identifying the specific reading skills necessary to students when responding to physics problems.

7. Conduct research relating the learning of students when models, demonstration apparatus, and other three-dimensional objects are used, to the aptitudes of the students.

BIBLIOGRAPHY

Books

Bennett, George K., Harold G. Seashore, and Alexander G. Wesman. Differential Aptitude Tests, Fourth Edition Manual, Forms L and M. New York: The Psychological Corporation, 1966.

Bruner, J. S. The Process of Education. New York: Vintage Books, 1960.

Buros, Oscar K. (ed.) The Sixth Mental Measurements Yearbook. Highland Park, Ill.: The Gryphon Press, 1965.

Cronbach, Lee J. Essentials of Psychological Testing. New York: Harper and Row Publishers, 1970.

Draper, N. R., and Smith, H. Applied Regression Analysis. New York: John Wiley and Sons Inc., 1968.

Ferguson, George A. Statistical Analysis in Psychology and Education. New York: McGraw-Hill Book Co. Inc., 1959 & 1966.

Guilford, J. P. Personality. New York: McGraw-Hill Book Co. Inc., 1959.

______________. Psychometric Methods. New York: McGraw-Hill Book Co. Inc., 1954.

Hukins, A. A. "A Factorial Investigation of Science Tests." Unpublished Doctoral Thesis. University of Alberta, Edmonton, 1963.

Jackson, Robert W. B., and Ferguson, George A. Studies in the Reliability of Tests, Bulletin Number 12, Department of Educational Research. Toronto: University of Toronto Press, 1941.

Kelly, Francis J., et al. Research Design in the Behavioral Sciences, Multiple Regression Approach. Carbondal and Edwardsville: Southern Illinois University Press, 1969.

Lord, Frederic M., and Novick, Melvin R. Statistical Theories of Mental Test Scores. Reading, Mass.: Addison-Wesley Publishing Co., 1968.

Roe, Anne. The Psychology of Occupations. New York: John Wiley and Sons Inc., 1966.

Sharo, E. A. "Physics, Mathematics and Visual Spatial Relations." Unpublished Doctoral Thesis. Columbia Teachers College, New York, 1962.

Smith, I. M. Spatial Ability. London: University of London Press, 1964.

Spearman, C. The Nature of Intelligence and the Principles of Cognition. London: MacMillan and Co. Ltd., 1927.

Thurstone, L. L., and T. G. Thurstone. Factorial Studies of Intelligence. Chicago: University of Chicago Press, 1941.

Thurstone, L. L. Primary Mental Abilities. Chicago: University of Chicago Press, 1938.

Vernon, Phillip E. The Structure of Human Abilities. London: Methuen and Co. Ltd., 1961.

Periodicals

Berdie, Robert F. The Differential Aptitude Tests as Predictors of Engineering Training. J. Ed. Psych., 1951, 42: 114-123.

Blade, M. F., and W. S. Watson. Increase in Spatial Visualization Test Scores During Engineering Study. Psych. Mono., 1955, Vol. LXIX, No. 12, 1-13.

Buell, R. R. Piagetian Theory into Inquiry Action. Sc. Ed., 1967, Vol. 51, 21-24.

Dempster, J. J. B. Symposium on the Selection of Pupils for Different Types of Secondary Schools. Br. J. Educ. Psych., 1948, Vol. XVIII, Part 3: 121-133.

Dressel, P. How the Individual Learns Science, Rethinking Science Education, Fifty-ninth Yearbook of the National Society for the Study of Education. Chicago: University of Chicago Press.

Fruchter, Benjamin. Measurement of Spatial Abilities. Ed. and Psych. Meas., 1954, Vol. XIV, Part 2: 387-395.

Lewis, D. G. The Factorial Nature of Attainment in Elementary Science. Br. J. Ed. Psych., 1964, 34: 1-9.

Michael, W. B., et al. Description of Spatial Visualization Abilities. Ed. and Psych. Meas., 1954, Vol. XIV: 185-199.

Mukerjhee, B. N. Factorial Structure of Aptitude Tests at Successive Grade Levels. Br. J. Stat. Psych., 1962, 15: 59-65.

Myers, Charles T. A Note on a Spatial Relations Pre-test and Post-test. Ed. and Psych. Meas., 1953, 13: 596-600.

Roe, Anne. The Psychology of the Scientist. Science, 1961, 134:456-9.

APPENDIX

APPENDIX A

TEST SCORES FOR THE SAMPLE OF STUDENTS

	Physics					D A T				
I.D. No.	PhV	PhN	PhG	PhDV	PhPD	VR	NA	AR	MR	SR
100	09	07	05	06	07	36	19	34	50	38
101	06	04	05	05	05	39	29	38	55	35
102	08	05	05	06	02	41	38	46	50	38
103	10	07	05	08	06	36	23	39	46	28
104	13	09	09	15	08	42	32	40	63	57
106	10	06	08	10	07	41	28	39	46	33
107	06	10	03	08	06	31	31	33	42	23
108	06	03	01	05	06	36	25	37	54	39
109	14	11	07	14	08	38	37	39	49	34
110	09	06	05	08	10	46	36	41	58	48
111	11	05	04	05	04	34	31	40	48	34
113	11	07	07	12	05	40	35	40	46	26
114	05	06	06	05	06	46	35	36	49	41
115	10	08	06	10	09	46	32	44	51	50
116	07	07	05	06	07	35	22	41	52	39
117	06	03	05	07	04	38	26	44	51	39
118	05	06	07	08	06	41	22	39	53	46
119	12	03	06	12	07	45	33	40	52	36
121	12	05	08	10	03	44	36	41	50	26
123	11	08	04	15	07	42	35	39	62	57
124	08	05	06	10	09	40	37	43	56	46
127	08	03	02	08	02	42	27	41	46	28
129	07	05	04	06	04	39	29	38	43	32
130	09	07	07	12	06	45	36	47	49	47
132	13	08	06	12	08	42	36	45	45	44
133	12	12	08	11	06	46	37	41	51	39
134	13	09	07	14	09	42	35	42	66	55
136	07	09	03	13	08	35	32	45	39	40

APPENDIX A (cont'd.)

	Physics					D A T				
I.D. No.	PhV	PhN	PhG	PhDV	PhPD	VR	NA	AR	MR	SR
137	07	07	05	10	02	41	37	45	54	52
200	07	04	03	04	06	36	30	43	51	38
201	08	06	05	10	08	42	28	38	45	33
203	06	07	06	11	04	38	32	41	56	37
204	12	10	07	14	08	41	27	41	55	43
206	09	04	03	11	05	34	31	39	45	37
207	09	07	07	06	07	43	31	37	51	30
208	13	09	07	11	08	47	38	45	54	42
210	13	11	08	14	09	48	34	43	65	49
212	09	05	06	08	06	40	36	41	53	45
214	09	06	04	07	06	39	31	45	47	28
215	07	04	05	09	07	34	26	45	56	44
216	07	05	00	05	09	33	34	42	47	37
217	12	04	05	06	07	41	35	43	49	24
218	13	11	08	14	09	46	39	47	66	54
221	12	09	07	13	08	50	31	46	62	48
224	13	09	07	13	07	45	36	46	59	55
226	09	05	08	10	04	30	36	37	49	25
227	04	03	02	08	02	25	18	41	46	49
229	08	08	03	09	09	41	37	45	60	52
231	11	09	06	11	06	42	37	46	59	36
232	07	03	04	06	06	27	23	35	54	34
301	07	05	03	12	05	36	30	34	52	36
302	14	11	08	11	09	40	39	43	58	56
303	12	09	04	11	07	45	33	44	63	48
304	03	02	02	06	05	38	30	37	43	27
306	12	08	07	11	07	43	40	47	59	54
308	10	07	04	03	05	38	31	37	38	36
309	08	06	06	08	05	36	36	39	50	37
310	06	05	04	05	07	26	35	37	55	40

APPENDIX A (cont'd.)

I.D. No.	PhV	PhN	PhG	PhDV	PhPD	VR	NA	AR	MR	SR
	Physics					D A T				
311	06	06	04	09	06	40	22	39	52	24
312	08	05	04	07	03	38	27	48	51	46
313	06	07	04	07	05	38	29	35	55	33
315	06	02	03	03	04	22	29	40	43	44
316	08	06	02	07	06	33	37	47	42	35
317	06	05	05	09	04	36	28	42	51	38
318	04	01	02	03	06	29	24	43	47	29
320	12	08	08	13	07	47	33	47	60	59
321	10	11	09	10	07	37	35	37	56	44
322	07	07	05	08	05	38	38	43	51	39
324	03	03	03	05	06	34	26	41	50	35
325	05	08	02	09	08	39	37	45	51	54
327	09	06	03	05	05	34	28	43	38	26
329	06	03	06	04	05	38	32	24	36	22
332	10	09	09	12	08	46	37	49	62	54
333	05	06	01	07	05	38	27	44	54	39

APPENDIX B

TABLE XII

ITEM ANALYSIS OF THE PHYSICS TESTS

I. Verbal PhV

Mean = 8.73
Variance = 8.01
Standard Deviation = 2.8
Reliability Coefficient (K-R Formula 20) = 0.62

Item Number	Level of Difficulty	Biserial Correlation
2	.74	.65
8	.86	.34
10	.64	.63
13	.28	.54
17	.84	.27
19	.55	.41
20	.72	.60
24	.30	.50
27	.40	.67
41	.68	.67
46	.77	.64
56	.35	.60
58	.47	.23
60	.58	.81
64	.56	.67
Mean	.58	.54

II. Numerical PhN

Mean = 6.43
Variance = 6.22
Standard Deviation = 2.5
Reliability Coefficient (K-R Formula 20) = 0.59

Item Number	Level of Difficulty	Biserial Correlation
1	.84	.40
7	.42	.65
12	.45	.42
14	.34	.57
18	.65	.50
28	.81	.58
39	.77	.65
40	.49	.58
45	.42	.78
53	.39	.70
55	.26	.70
59	.61	.52
Mean	.54	.60

III. Graphic PhG

Mean = 5.11
Variance = 4.47
Standard Deviation = 2.1
Reliability Coefficient (K-R Formula 20) = 0.50

Item Number	Level of Difficulty	Biserial Correlation
4	.76	.49
5	.58	.63
6	.64	.64
25	.70	.71
26	.46	.73
33	.34	.74
34	.32	.52
35	.54	.51
42	.50	.48
43	.27	.27
Mean	.51	.55

IV. Diagrammatic-verbal PhDV

Mean = 8.86
Variance = 10.1
Standard Deviation = 3.2
Reliability Coefficient (K-R Formula 20) = 0.68

Item Number	Level of Difficulty	Biserial Correlation
11	.85	.56
15	.89	.25
21	.12	.53
23	.39	.44
26	.77	.61
37	.73	.83
38	.68	.78
44	.40	.52
47	.18	.67
48	.64	.51
49	.51	.44
50	.58	.77
51	.43	.65
52	.13	.34
61	.65	.59
62	.50	.52
63	.42	.45
Mean	.53	.56

V. Pure Diagrammatic PhPD

Mean = 6.19
Variance = 3.53
Standard Deviation = 1.8
Reliability Coefficient (K-R Formula 20) = 0.45

Item Number	Level of Difficulty	Biserial Correlation
3	.28	.47
9	.46	.66
16	.30	.41
29	.74	.75
30	.90	.74
31	.64	.58
32	.77	.81
54	.74	.39
57	.72	.62
65	.64	.40
Mean	.62	.58

APPENDIX C

GRADE XI PSSC TESTS

Answer the following test items on the answer sheet provided with this test. Select the best answer from the five possible answers for each question. Blacken the space which corresponds to the correct answer by using the pencil provided.

Note: Blacken the space between the dotted lines completely so that it forms a clear black mark - do not go over the lines.

Do not spend too much time on any one item. If you cannot do a question go on to the next one. Work as quickly and accurately as you can.

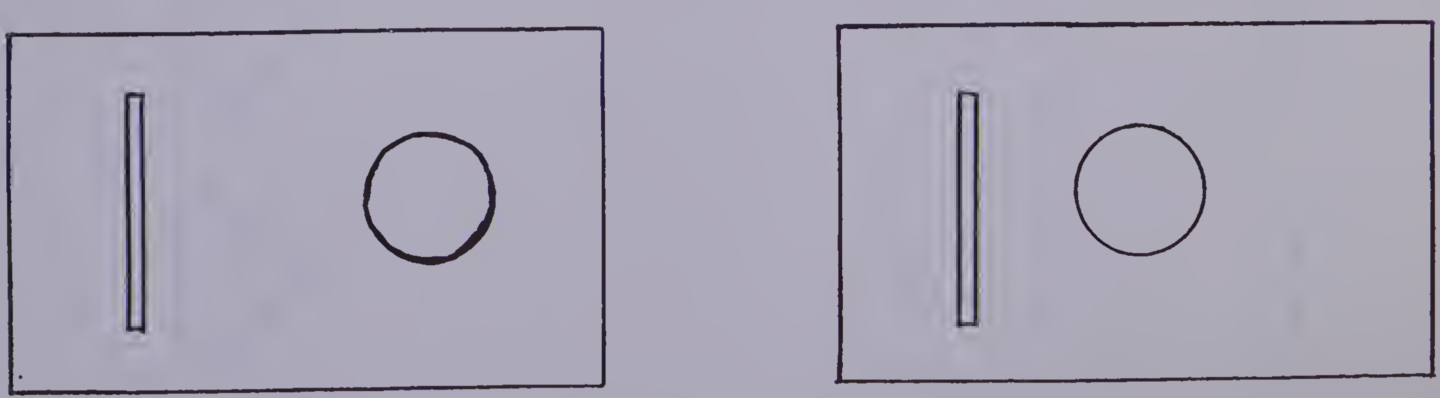
1. A strobe disc with 6 openings "stops" the motion of a rotating wheel. If the disc turns 22 times in 10 seconds then the speed of rotation of the wheel is:

   a) 2.2 rev./sec.
   b) 1.3×10^3 rev./sec.
   c) 3.7×10 rev./sec.
   d) 2.8×10^2 rev./sec.
   e) 1.3×10 rev./sec.

2. The process of extending a graph beyond the range of the data we have collected is called:

   a) Interpolation
   b) Interpretation
   c) Scaling up
   d) Extrapolation
   e) Variation

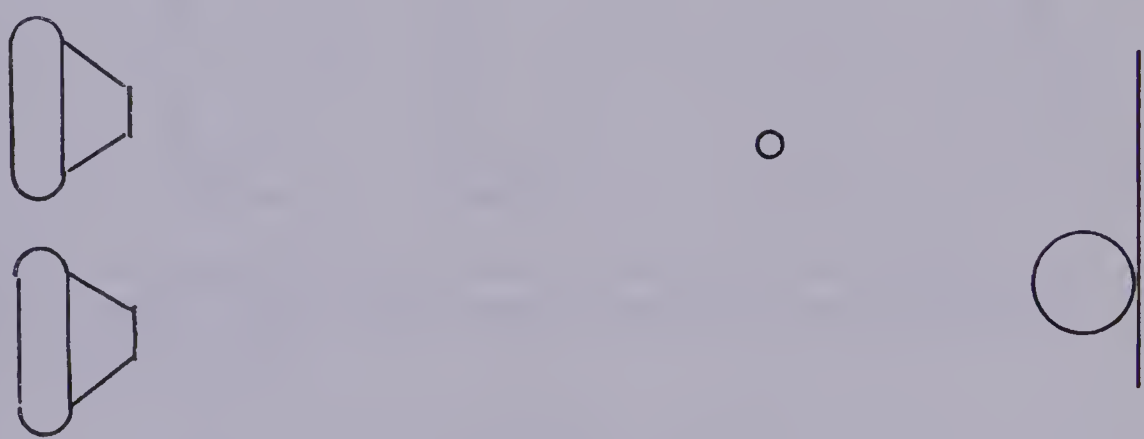
3. Two cameras take pictures shown below of a rod and a ball.



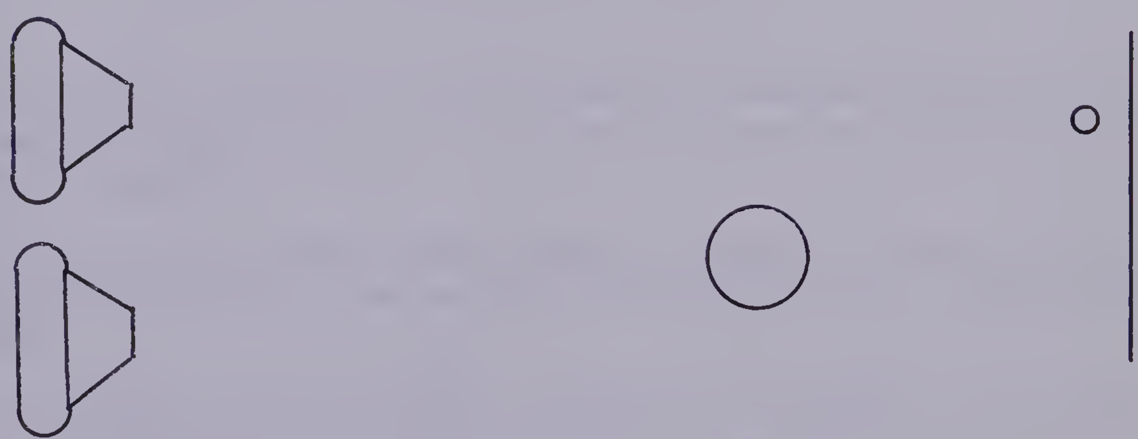
Choose the correct arrangement of the cameras; ball and rod.

a)

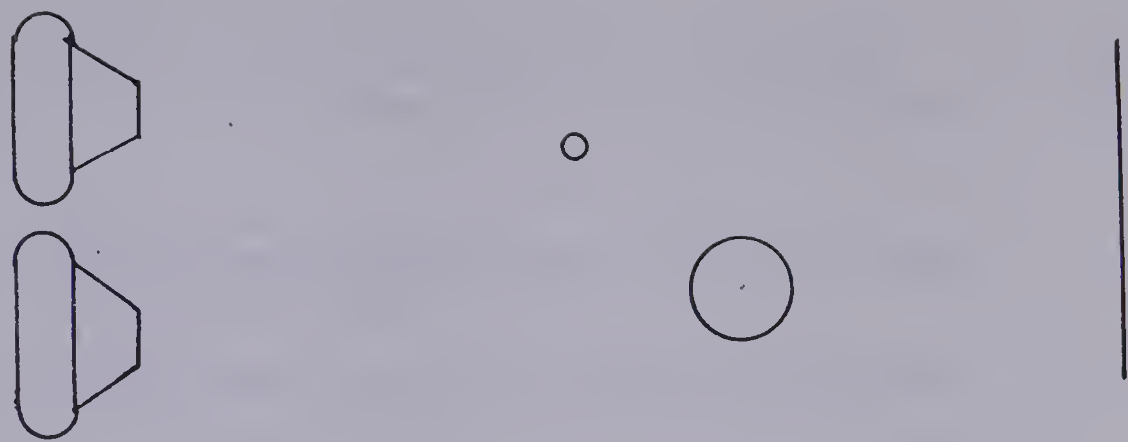
b)



c)



d)

e) 

Questions 4, 5 and 6 refer to the following graphs.

4. If the horizontal axis is the x axis and the vertical axis is the y axis, which graph best represents the relationship between the following data?

X	3	6	8	4	2	1	10
Y	7.5	15	20	10	5	2.5	25

5. Which graph would show the relationship between the radius of a circle (on the horizontal axis) and the area of a circle (on the vertical axis)?

6. Which graph shows the relationship in which the value of Y (vertical axis) is constant for every value of X (horizontal axis)?

7. The contents of a large bin of grain is required. It is stored in a cylindrical granary 24.0 feet in diameter. Its height is 9 ft. The volume of 1 bushel of grain is 1.50 cubic feet. The number of bushels contained by the granary is best stated to the correct number of significant figures, as:

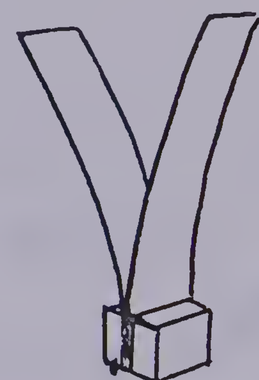
a) 452
b) 5×10^2
c) 1×10^2
d) 3×10^3
e) 2713

8. In "stopping" certain kinds of motion which are too fast to be seen with the unaided eye we may use a stroboscope. The stroboscope is useful for stopping only one of the following. Which one:

a) A rifle bullet.
b) A single wave travelling along a stretched string.
c) An airplane propeller.
d) A light ray.
e) The shock wave in front of a supersonic airplane.

9. 15 revolutions in 12 seconds

Makes 60 vibrations in 12 sec.



The observer sees:

a)

b)

c)

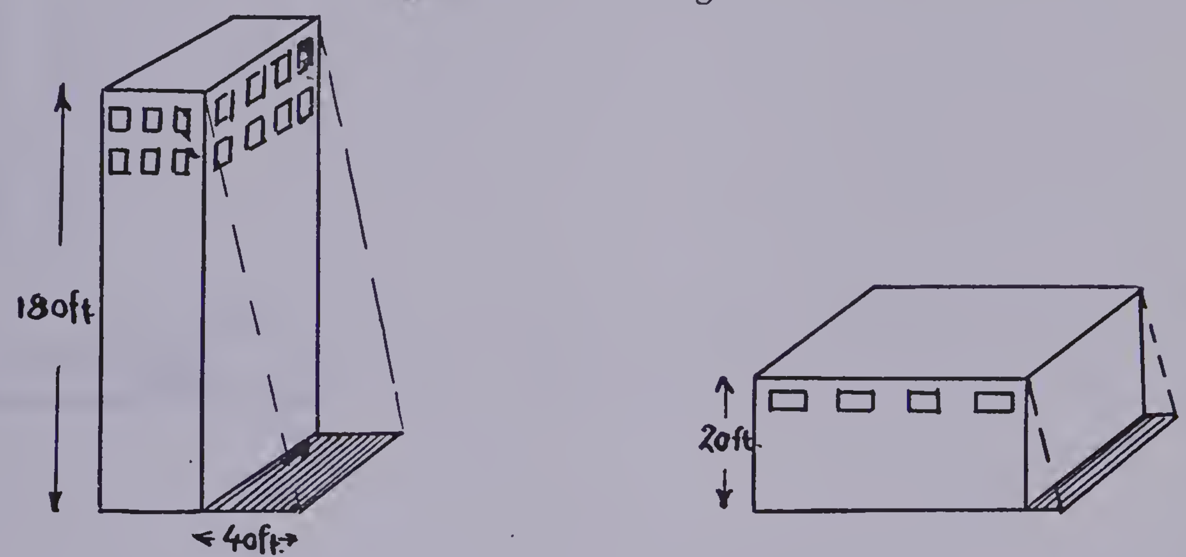
d)

e)

10. The largest base line available to astronomers for the measurement of distances to the stars, is:

a) The measured mile.
b) The circumference of the earth.
c) The circumference of the earth's orbit.
d) The diameter of the earth.
e) The diameter of the earth's orbit.

11. From the shadows cast by the buildings shown below, the length of the shadow of the smaller building is:



a) 360 feet
b) 90 feet
c) 9 feet
d) 4.4 feet
e) .44 feet

12. A boy sees the full moon and holds a ruler two feet from his eye. The moon appears to have a width of .25 inches as measured by the ruler 2.0 feet from his eye. He knows that the moon is about 250,000 miles from the earth. He finds it has a diameter of:

a) 2.4×10^7 miles

b) 2.6×10^3 miles

c) 1.2×10^5 miles

d) 1.5×10^4 miles

e) 3.1×10^4 miles

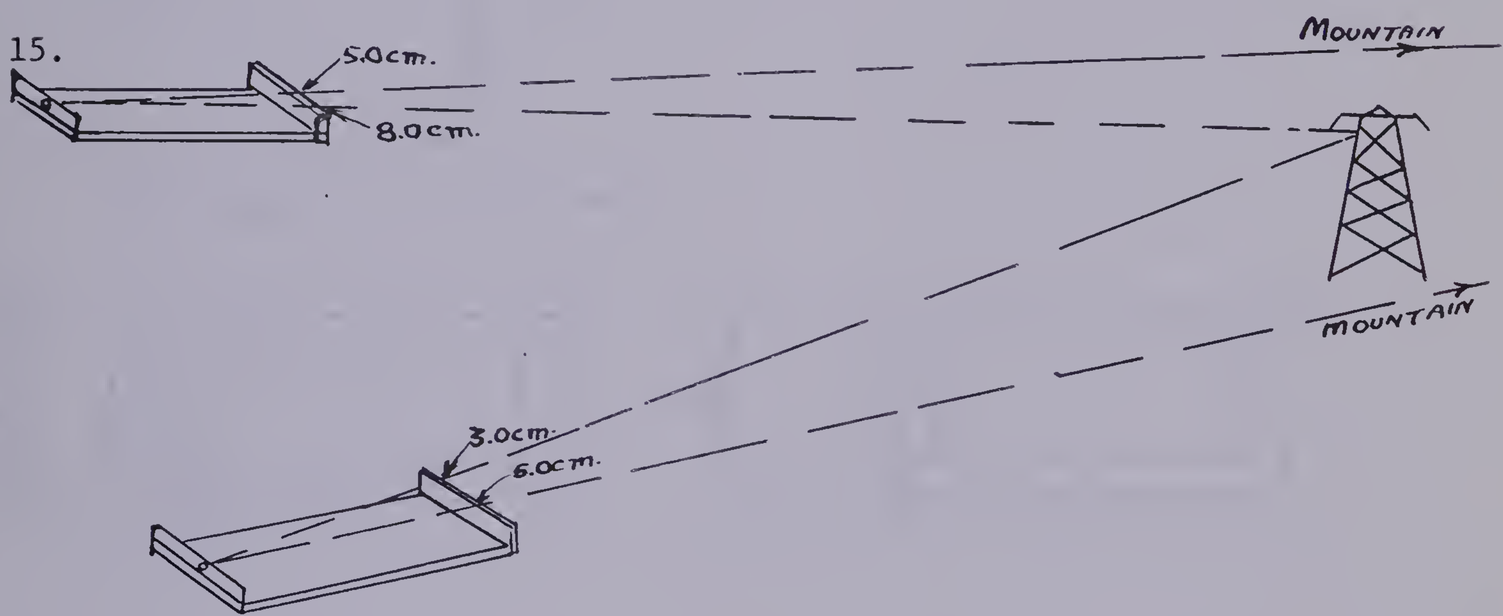
13. An example of two quantities which are inversely related is:

a) The volume of a sphere and its radius.
b) The time it takes to travel 100 miles and the speed.
c) The distance travelled by a car at constant speed and the time.
d) The volume of a sphere and its radius.
e) The strength of a rope and its length.

14. The order of magnitude of the number of ping pong balls (diameter 1") required to fill a room 30 ft. by 60 ft. by 8 ft. is:

a) 10^5
b) 10^4
c) 10^8
d) 10^6
e) 10^7

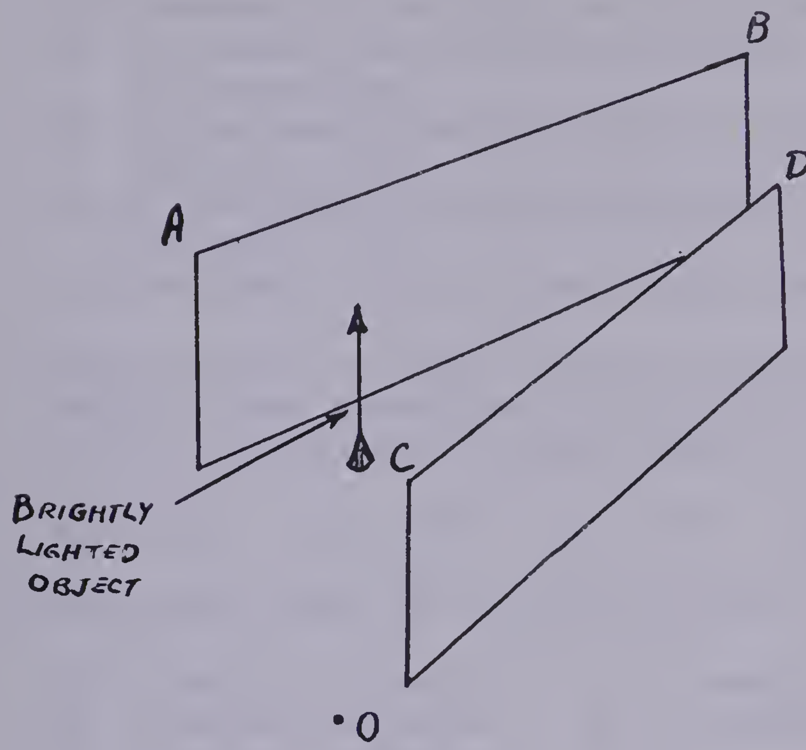
15.



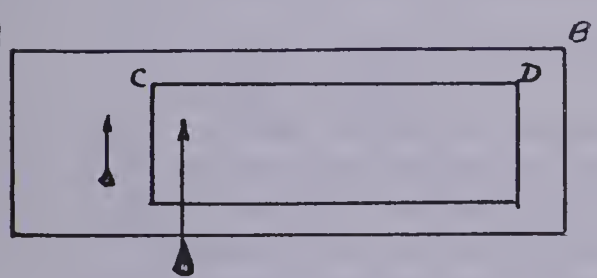
A parallax viewer is used to determine the distance of tower X. The tower is sighted at the 8.0 cm. mark of the viewer while a very distant mountain is in line with the center of the viewer (5.0 cm. mark). The observer then moves 15 meters along a line perpendicular to the direction of the mountain. The tower is again sighted, this time at the 3.0 cm. mark of the viewer while the mountain remains in line with the 5.0 cm. mark. If the viewer is 50 cm. long, the distance to tower X is:

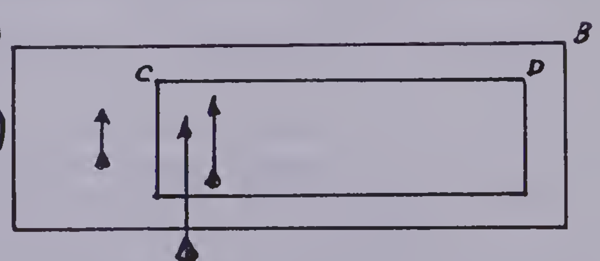
a) 150 meters
b) 3750 meters
c) 68 meters
d) 250 meters
e) 94 meters

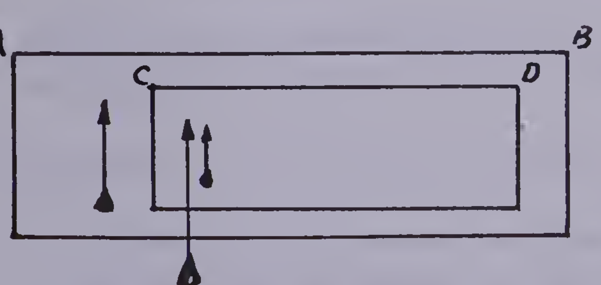
16. AB and CD are plane mirrors, the silvered sides face one another.



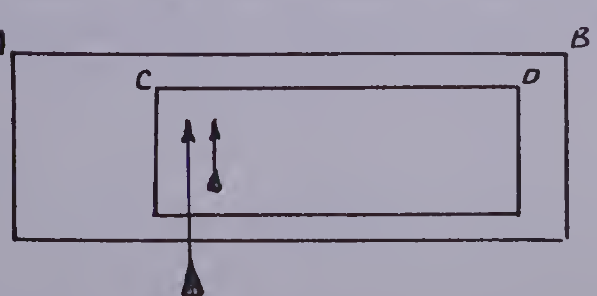
The observer at O will see:

a) 

d) 

b) 

e) 

c) 

17. The Canadian flag as seen in pure green light would appear:

a) Completely green with very little pattern observable.
b) A red maple leaf and end bars with a green background for the maple leaf.
c) A green maple leaf and end bars with a dark background for the maple leaf.
d) A dark maple leaf and end bars with a green background for the maple leaf.
e) Completely dark with very little pattern observable.

18. The critical angle for a pencil of light passing from glass ($n=1.60$) into plastic is 63^{o} ($\sin 63 = .891$). The absolute index of refraction for the plastic is:

a) 2.23 b) 1.79 c) 2.49 d) 1.35 e) 1.43

19. Pure white paper is best described as:

a) A regular reflector b) A specular reflector
c) A normal reflector d) A diffuse reflector
e) A partial reflector

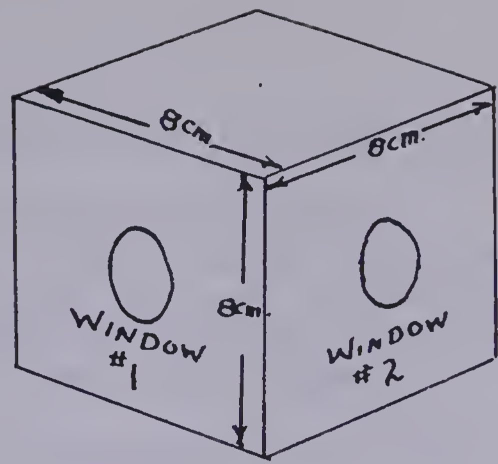
20. One significant distinction between real and virtual images is:

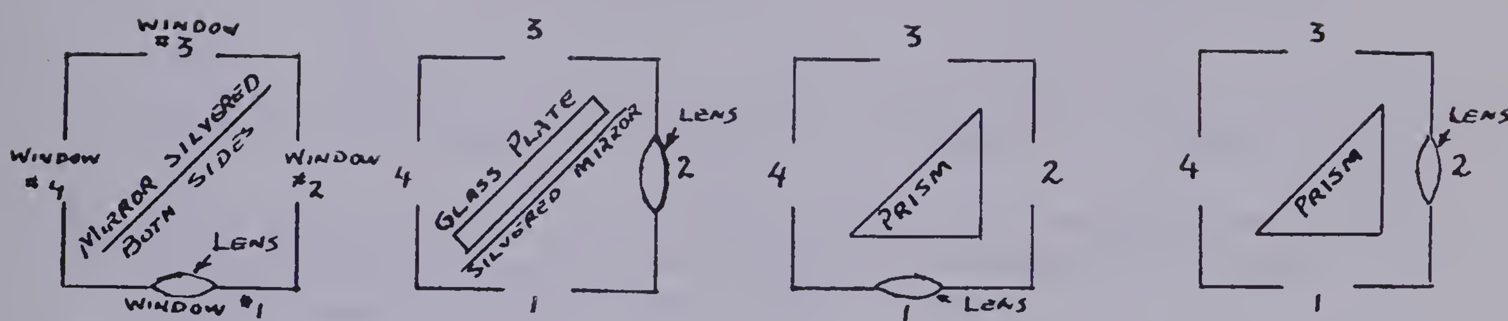
a) A real image can be displayed on a screen and a virtual image cannot.
b) A real image is larger than a virtual image.
c) A virtual image is larger than a real image.
d) A virtual image is produced by mirrors only while a real image may be produced by either mirrors or lenses.
e) A real image is formed by refraction, a virtual one by reflection.

ANSWER THE FOLLOWING QUESTIONS BY SELECTING THE BEST ANSWER FROM AMONG THE POSSIBLE ANSWERS GIVEN AT THE END OF THE QUESTION.

Questions 21, 22 and 23 relate to the following diagrams and information.

A cubical box, 8 cm. on an edge, has four small windows, one on each vertical side as shown in the diagram. The windows are numbered 1, 2, 3, and 4. An observer notices that when he looks into window 4, he sees a virtual image of an object that is outside the box near window 3. He suggests several possible models of the box to explain these observations. The diagrams below show these models.





Model I　　Model II　　Model III　　Model IV

21. Which of the models will satisfactorily explain his observation?

a) I and II only
b) III and IV only
c) I and III only
d) II and IV only
e) I, II, III, IV.

22. The observer places an object 5.0 cm. outside window 2. When he holds a piece of paper 3.0 cm. outside window 1, he finds a sharp image of the object is formed on the paper and that it is smaller than the object. On the basis of this observation and of the initial observation the satisfactory models are:

a) I and II only
b) III and IV only
c) I and III only
d) II and IV only
e) I, II, III, and IV

23. As a further check, the observer makes a small hole through the top of the box and fills the box with a liquid of index of refraction 1.5. He then looks through window 1 and notices that he can see right through the box and out window 3, and when he looks through window 2 he can see straight through the box to window 4.

Which of the originally proposed models is satisfactory on the basis of all of the observations?

a) I and II
b) I and III
c) I only
d) III only
e) IV only

24. The statement i/r = index of refraction is only true, if:

a) i and r are small angles.
b) i and r refer to light travelling from air to water.
c) i and r are measured in materials of small index of refraction.
d) the light is travelling from a vacuum into another medium.
e) i and r are large angles.

Questions 25 and 26 are related to the following graphs:

25. As a man walks away from a plane mirror at a uniform rate (velocity), which graph most nearly represents the distance of his image from him plotted against time?

26. As a man walks toward a plane mirror at uniform velocity, which graph most nearly represents the distance of his image from him plotted against time?

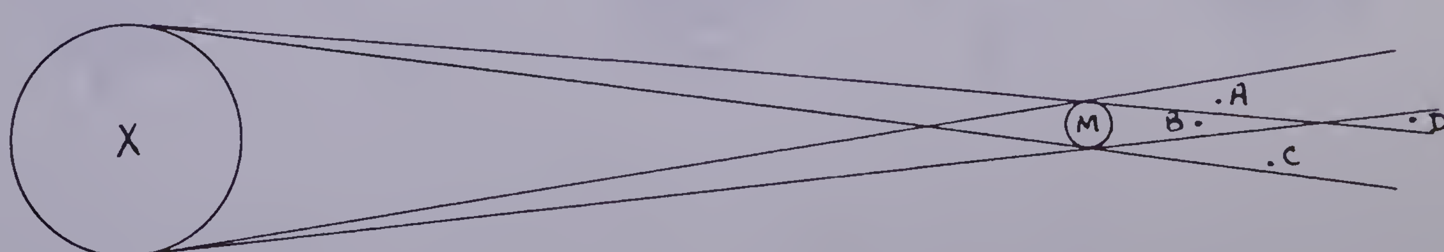
27. If one looks at a long, thin source of light through a slit which is about 1/10 mm. wide, what one sees is a series of alternating light and dark bands. This phenomenon is called:

a) Refraction
b) Penumbra
c) Dispersion
d) Diffraction
e) Deviation

28. A narrow pencil of light passes from a vacuum into glass (index of refraction = 1.50). If the sine of the angle in the glass is 0.400, what will be the sine of the angle in the vacuum?

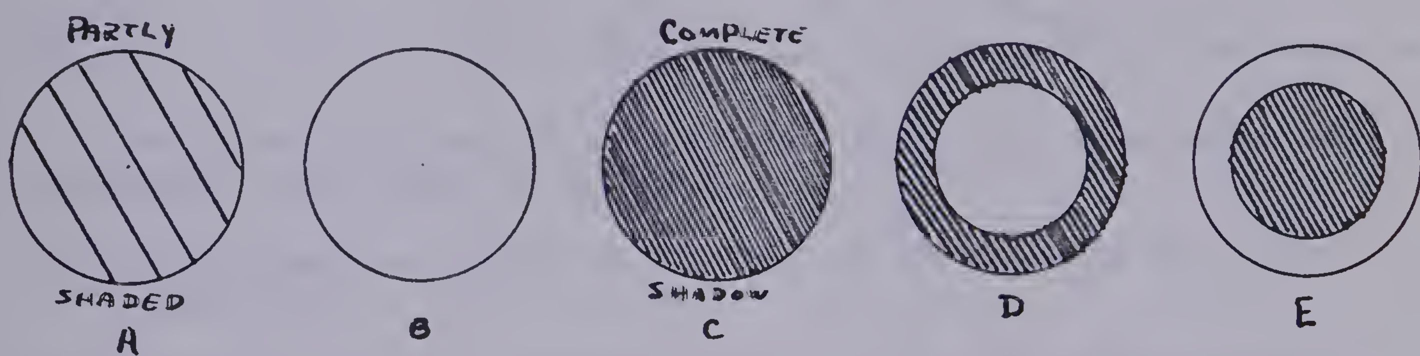
a) 0.60 b) 0.55 c) 0.25 d) 0.27 e) 1.90

In the following diagram X represents the sun, M represents the moon. The positions indicated by the letters A, B, C, and D are positions of the observer in space.

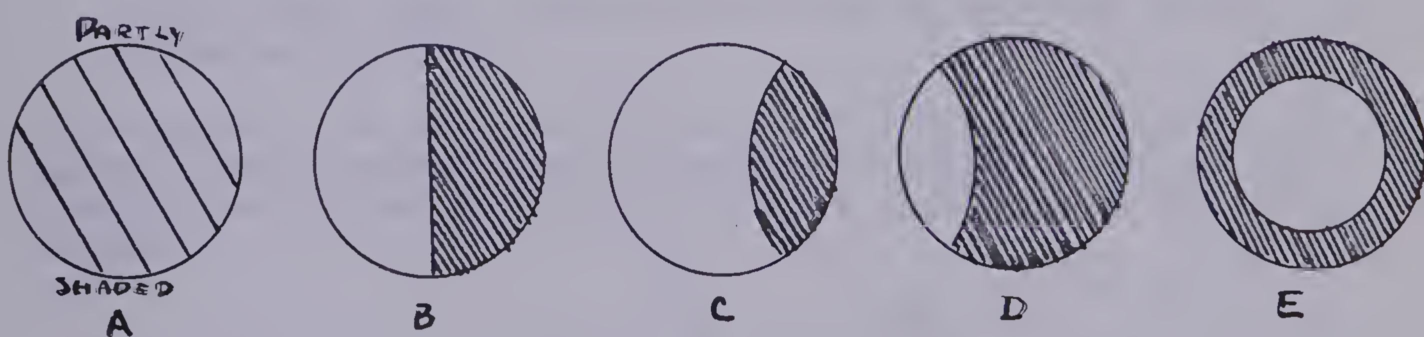


29. What would an observer see from Position A?

30. Position B?

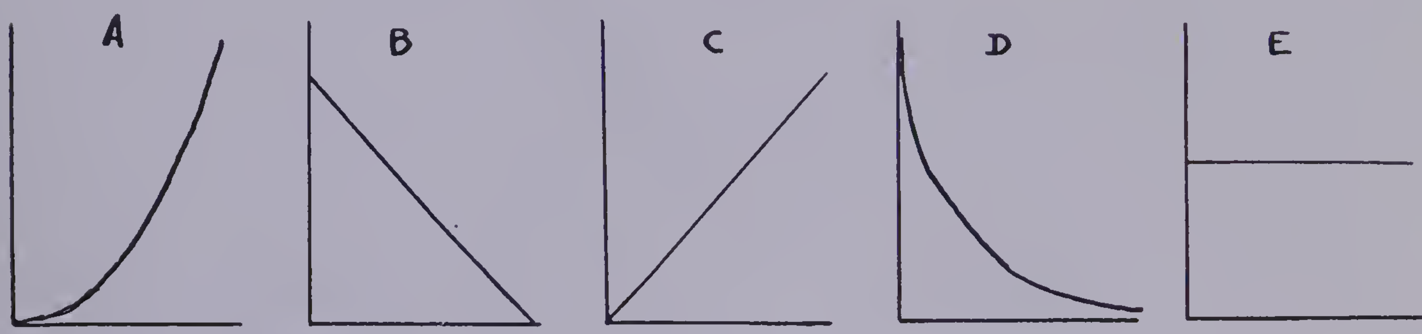


31. Position C?



32. Position D?

Questions 33 to 35 relate to an optical experiment. Each question is to be answered with one of the five graphs below. The horizontal axis represents the object distance. A graph my be used once, not at all, or more than once.



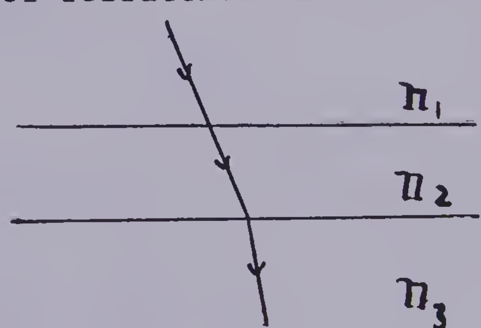
An object is placed at the principal focus of a converging lens and moved slowly away from the lens.

33. Which graph represents the location of the image as the object moves away?

34. Which graph could represent the product of the object distance times the image distance as the object moves away?

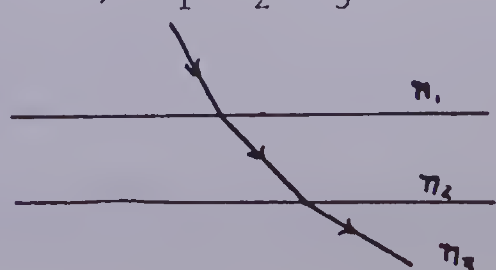
35. Which graph could represent the height of the image as the object moves away?

The diagrams below show the path of a narrow pencil of light as it passes through three different transparent materials. Select the combination of indexes of refraction which would account for each diagram.



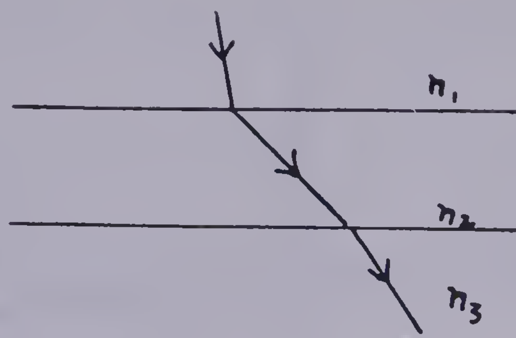
36. a) $n_1 > n_2 > n_3$ b) $n_1 > n_2 = n_3$ c) $n_1 = n_2 > n_3$

d) $n_1 = n_2 < n_3$ e) $n_1 = n_2 = n_3$



37. a) $n_1 > n_2 > n_3$ b) $n_1 > n_3 > n_2$ c) $n_3 > n_2 > n_1$

d) $n_2 > n_3 > n_1$ e) $n_3 = n_1 > n_2$



38. a) $n_1 > n_2 > n_3$ b) $n_3 > n_2 > n_1$ c) $n_1 > n_3 > n_2$

d) $n_2 > n_3 > n_1$ e) $n_1 = n_3 > n_2$

39. A narrow pencil of light travels from glass (n = 1.50) into a liquid. If the angle in the glass is 30° (sin 30 = .500) and the sine of the angle in the liquid is .600, what is the absolute index of refraction for the liquid?

a) 1.40 b) 1.25 c) 0.80 d) 1.20 e) 0.30

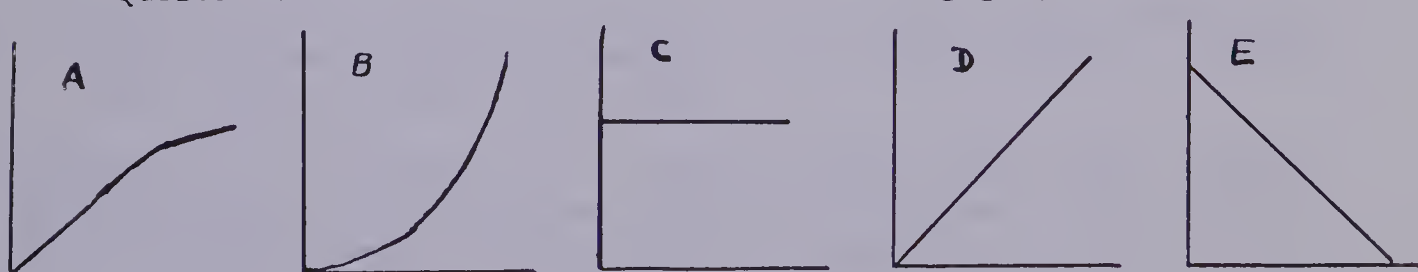
40. A transparent material has a critical angle equal to 40°. Its index of refraction will be: Sin 40° = .643.

a) 0.64 b) 1.64 c) 1.56 d) 1.36 e) 1.33

41. A material such as greased paper which scatters light and also transmits it is said to be:

a) Translucent b) Opaque c) Absorptive
d) Transparent e) Incandescent

Questions 42 and 43 relate to the following graphs.

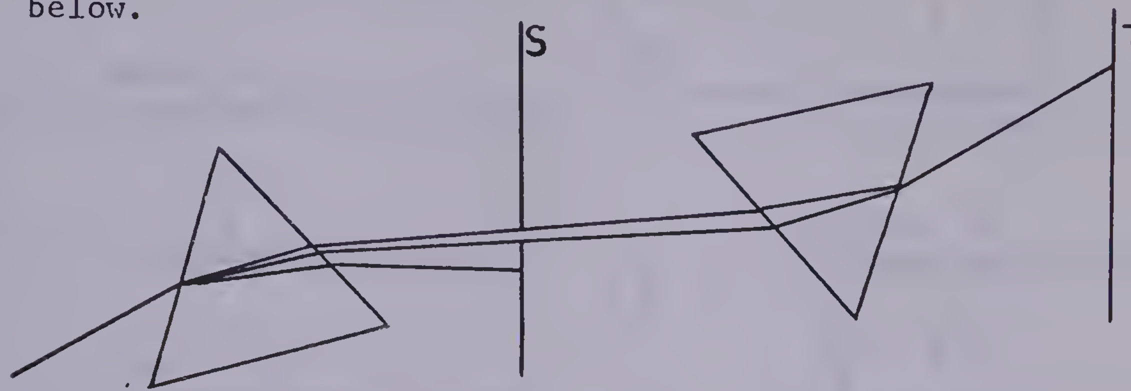


42. If a pencil of light travels from air to glass, which of the graphs would represent the relationship between sin i and sin r. Where sin i is plotted along the horizontal axis.

43. If a pencil of light travels from air to glass (Index of refraction = 1.50) which graph could represent the relationship

between i and r where i is measured along the horizontal axis and r is measured along the vertical axis?

44. A beam of white light passes through the set of prisms shown below.



The Screen S is arranged so that only the part of the spectrum shown above is allowed to continue on the second prism. On the screen marked T what will you see?

a) a spot of red light
b) a spot of white light
c) a dark spot - no light gets through
d) a green spot
e) a violet spot

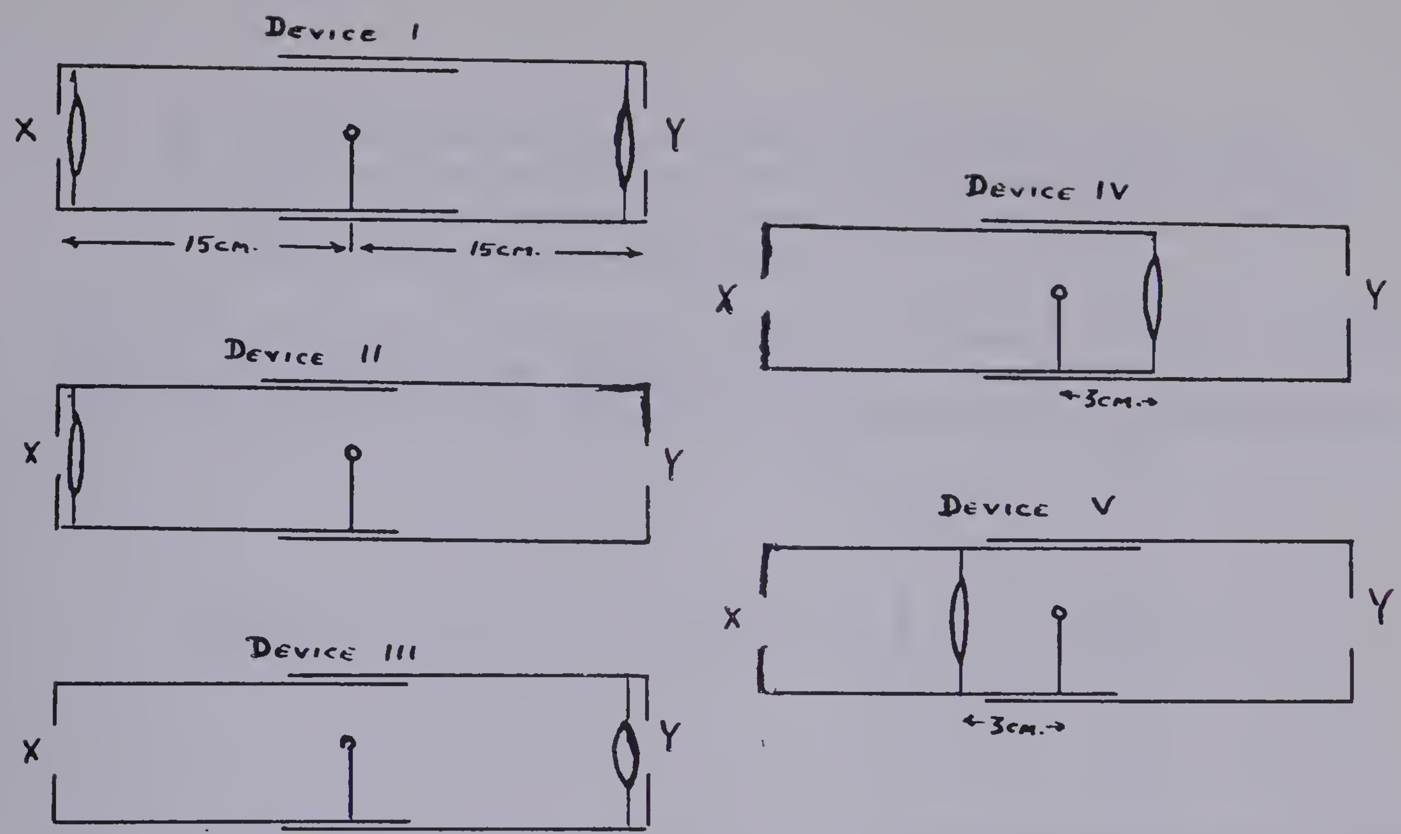
45. An object 4 cm. tall is placed 10 cm. in front of a convex lens whose focal length is 6 cm. The image formed will be:

a) virtual, magnification = 0.6, Si = 3.6 cm.
b) real, magnification = 0.56, Si = 2.25 cm.
c) virtual, magnification = 1.5, Si = 9 cm.
d) real, magnification = 0.6, Si = 3.6 cm.
e) real, magnification = 1.5, Si = 9 cm.

46. When an object is placed closer to a parabolic reflector than the focus, the image form is:

a) Virtual, erect, enlarged.
b) Real, erect, smaller.
c) Virtual, inverted, smaller.
d) Real, inverted, smaller.
e) Virtual, erect, same size.

Questions 47 - 50 refer to the optical devices shown below. Each device contains an illuminated vertical pin and one or two lenses (focal length = 5 cm.). All lenses are identical.



47. When you look through X, the pin appears erect; when you look through Y, it also appears erect but smaller than it appears through X. Which device is it?

a) I b) II c) III d) IV e) V

48. When you look through X, the pin appears inverted; when you look through Y it appears erect. Which device is it?

a) I b) II c) III d) IV e) V

49. You push the two ends of the device together, using the sliding joint. When you look through X, the pin always appears inverted and always the same size. When you look through Y, the pin appears first to be inverted, then it becomes blurred, and finally appears erect. Which device is it?

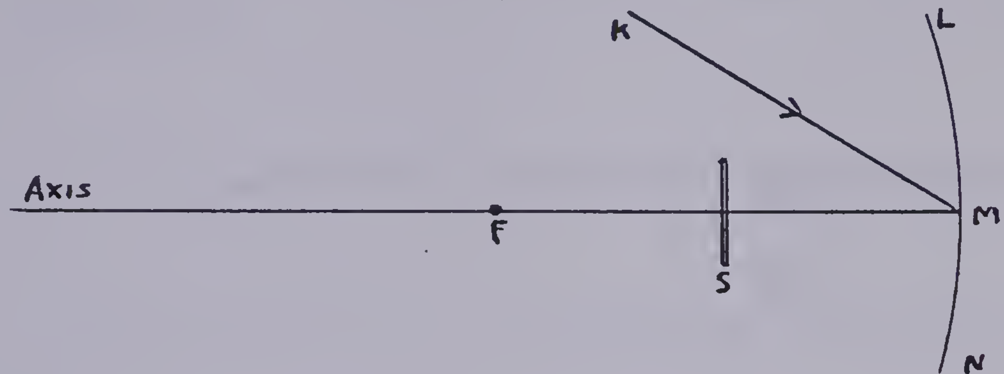
a) I b) II c) III d) IV e) V

50. Device IV is completely filled (to the openings) with a transparent colorless material whose index of refraction is equal to that of the lens. Which of the following effects would be observed?

a) The focal length of the lens is decreased.
b) The image of the pin as seen through end Y would be inverted.

c) The pin appears the same when seen through either end.
d) The image of the pin as through end Y would appear larger.
e) The pin would appear brighter when viewed through end X than when viewed through end Y.

Questions 51 and 52 refer to the parabolic mirror L M N whose principal focus is F. This mirror is shown below.
M is the midpoint of the mirror, at which the axis touches the mirror.



51. KM represents a ray of light striking the parabolic mirror at its midpoint. This light will be reflected.

a) along the line which passes above K and below L.
b) along the line M K.
c) along the line M F.
d) along a line which passes below F and above N.
e) according to Snell's Law.

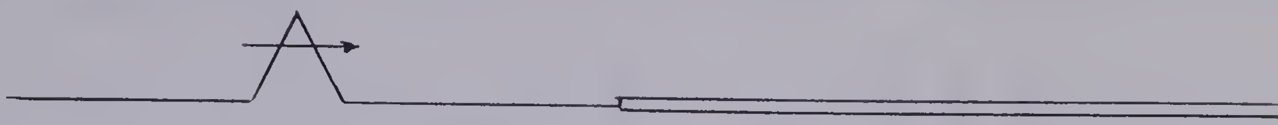
52. The mirror is turned so that its axis points to a distant star. S represents a small mirror located halfway between F and M with its reflecting surface facing M and perpendicular to the axis of the parabolic mirror. Which of the following statements most nearly describes the effect produced by the mirror S?

a) A real image of the star will be formed between S and M.
b) A real image of the star would be formed at M.
c) A real image of the star would be formed to the right of M if the surface of the parabolic mirror were not in the way.
d) No real image of the star will be formed because a plane mirror can form only virtual images.
e) No real image of the star will be formed because the plane mirror blocks off the light from the parabolic mirror.

53. A pencil of light passes from water (n = 1.33) into a transparent plastic block. The angle of incidence is 48° and the angle of refraction is 36°. The relative index of refraction from water to plastic is. Sin 48° = .743 Sin 36° = .588

a) 1.26 b) 1.78 c) 1.33 d) 1.68 e) 1.06

54. Medium 1 is a light spring - medium 2 is a heavy spring.



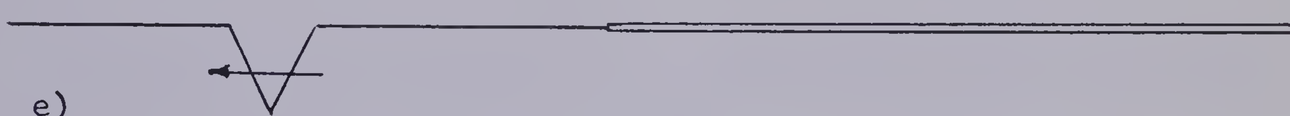
Which of the following shows what the pulse looks like after transmission?

a)

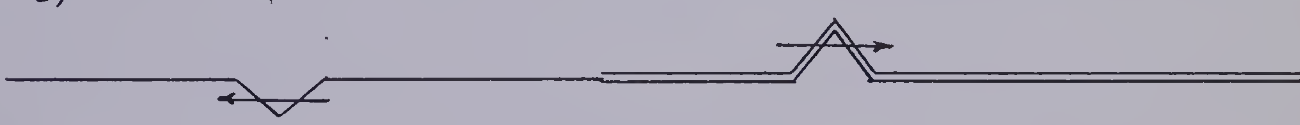
b)

c)

d)



e)



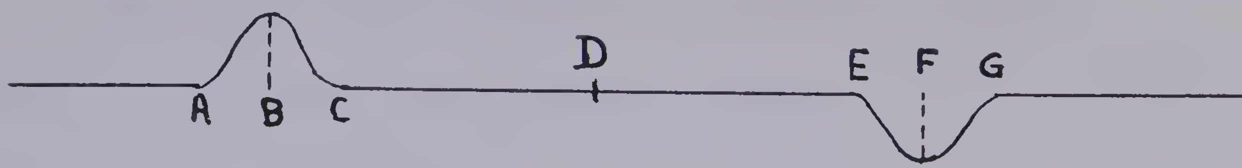
55. Two point sources of light are arranged on either side of a screen. The source on the left side is 2 feet away from the screen and the light source on the right side is 3 feet from the screen. If the illuminations from both sources at the screen are equal then the left hand source is how many times as bright as the right hand source?

a) 1.50 b) 1.00 c) 0.67 d) 0.44 e) 2.25

56. Which of the following properties of light does the particle model NOT explain satisfactorily?

a) Light pressure
b) Straight line propagation of light
c) The variation of the intensity of light with distance
d) The diffraction of light by narrow slits
e) The reflection of light by polished metals

57. Two waves in a spring travel toward one another as shown in the diagram below:

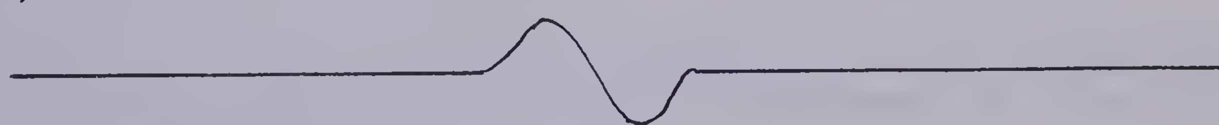


When the two waves overlap so that points E and B coincide, and points C and F coincide, the shape of the combined wave at that point will be most like which of the following?

a)


b)


c)


d)


e)


58. Waves which are produced at regular time intervals are said to be:

a) Periodic
b) Straight
c) Transverse
d) Refracted
e) Pulses

59. According to the particle model for light, the speed of light in a medium whose index of refraction is 1.60 would be:
Speed of light in air = 3.00×10^8 m./sec.

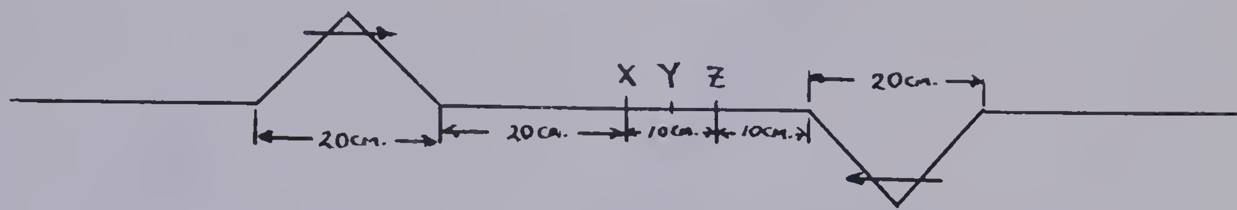
a) 3.00×10^8 m./sec.
b) 2.23×10^8 m./sec.
c) 4.80×10^8 m./sec
d) 1.87×10^8 m./sec.
e) 4.60×10^8 m./sec.

60. When light strikes a glass surface from air, it is observed to be partly refracted and partly reflected. The amount of light which is reflected:

a) Is small and constant for all angles of incidence.
b) Is large when the angle of incidence is small.
c) Is largest when the angle of incidence is zero.

d) Occurs only when the critical angle for a ray passing from air to glass is exceeded.
e) Is increased when the angle of incidence is large.

Questions 61 - 63 refer to the diagram below:



Two triangular-shaped pulses, identical in shape but opposite in displacement, travel toward each other along a rope. The length of the displaced portion of each pulse is 20 cm. and the points of maximum displacement, M and N, are each 10 cm. from the leading edge of their respective pulses. M and N, are each displaced 4.0 cm. from the rest position of the rope.

61. Point X is midway between the pulses. As the pulses pass through X, what is the maximum displacement X will undergo?

a) zero cm. b) 2.0 cm. c) 4.0 cm.
d) 6.0 cm. e) 8.0 cm.

62. Point Y is 5 cm. to the right of point X. When the leading edges of the two pulses coincide at X, the displacement of Y from its rest position will be most nearly

a) zero cm. b) 2.0 cm. downward c) 4.0 cm. upward
d) 2.0 cm. upward e) 4.0 cm. downward

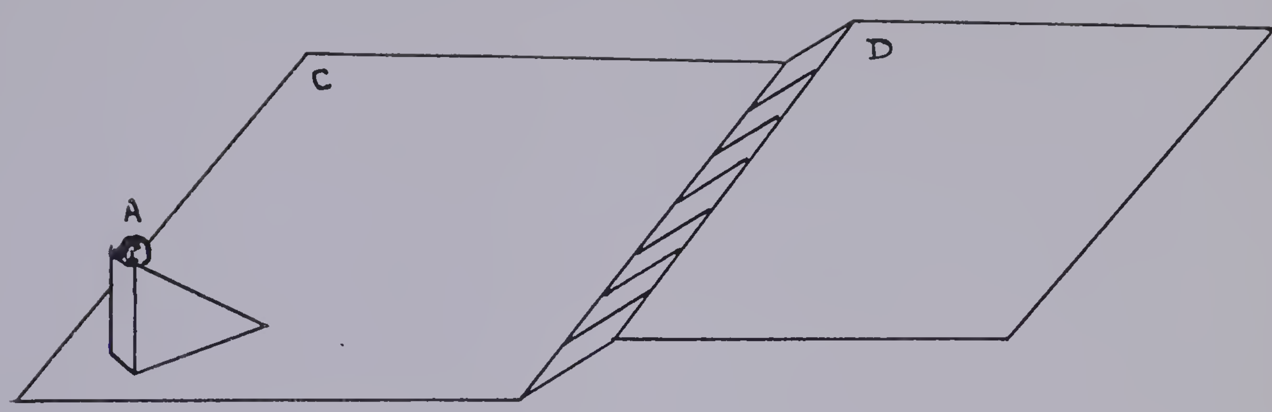
63. What will be the maximum displacement which point Y will undergo during the passage of these two pulses?

a) zero cm. b) 2.0 cm. c) 4.0 cm.
d) 6.0 cm. e) 8.0 cm.

64. Total internal reflection can be observed under which of the following conditions?

a) At a polished metal surface.
b) When light is transmitted from air to glass at angles bigger than the critical angle.
c) When light passes normally from air to glass or from glass to air.
d) When light is transmitted from glass to air at angles bigger than the critical angle.
e) When light enters glass from air at angles of incidence almost equal to 90°.

65.



C and D are both level surfaces. A ball on the inclined ramp A is given enough speed to climb up onto level D. Which diagram shows the path of the ball?